

STIC-Biotech/ChemLib

80934

From: Collins, Cynthia  
Sent: Friday, November 22, 2002 2:23 PM  
To: STIC-Biotech/ChemLib  
Subject: sequence search request SN 09/868300

Please search, both prior art and interference, for SN 09/868300:

- 1) SEQ ID NO:7
- 1) SEQ ID NO:8

Thank You,

Cynthia Collins  
Art Unit 1638  
CM1, 9A12 (office) or 9E12 (mailbox)  
(703) 605-1210

CRFE

12/17/98

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 12/4/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

Point of Contact  
P. Sheppard

Telephone number: (703) 308-4499

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2002, 19:10:05 ; Search time 2417 Seconds  
(without alignments)  
4375.529 Million cell updates/sec

Title: US-09-868-300-7  
Perfect score: 653  
Sequence: 1 gaatcgacacagagctccctt.....caaagagatgacacagcttg 653

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estbda:\*  
2: em\_esthum:\*  
3: em\_estlm:\*  
4: em\_estlm:\*  
5: em\_estlm:\*  
6: em\_estlm:\*  
7: em\_estlm:\*  
8: em\_estlm:\*  
9: gb\_est1:\*  
10: gb\_est1:\*  
11: gb\_est1:\*  
12: gb\_est1:\*  
13: gb\_est1:\*  
14: gb\_est1:\*  
15: em\_estlm:\*  
16: em\_estlm:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_hum:\*  
20: em\_gss\_hum:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_vit:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501.4	76.8	529	10	AM004542 701932172
2	345.6	52.9	536	9	A1994093 701498662
3	345.2	52.9	765	13	B1204061 EST522101
4	335.6	51.4	712	12	BG596323 EST495001
5	326.6	50.0	823	10	BE051140 GA_Fa003
6	322.2	49.3	745	10	AM218763 EST301243

7	320.8	49.1	557	9	AU236886	AU236886
8	314.2	48.1	566	14	BQ273330	BQ273330
9	311.6	47.7	441	17	BH634498	BH634498
10	311	47.6	452	14	R30079	R30079
11	276.2	42.3	1165	11	AY105313	AY105313
12	271.6	41.6	641	12	BG097619	BG097619
13	269.8	41.3	485	13	BM524990	BM524990
14	261.2	40.0	944	13	B1948508	B1948508
15	259	39.7	502	10	BE357810	BE357810
16	257	39.4	659	12	BG366429	BG366429
17	255.6	39.1	678	9	AL750043	AL750043
18	254.6	39.0	608	14	BQ415563	BQ415563
19	240.2	36.8	859	12	BF628880	BF628880
20	234.2	35.9	496	12	BG417248	BG417248
21	232.8	35.7	507	10	BE440848	BE440848
22	227.4	34.8	419	14	BQ273331	BQ273331
23	225.8	34.6	612	10	AM218764	AM218764
24	219	33.5	664	14	BH080391	BH080391
25	218	33.4	523	13	BM097653	BM097653
26	213.8	32.7	540	9	AJ470478	AJ470478
27	210.8	32.3	477	14	BQ753916	BQ753916
28	202.4	31.0	668	13	B1959690	B1959690
29	197	30.2	631	13	B1168895	B1168895
30	196.4	30.1	678	17	BH684219	BH684219
31	193.4	29.6	614	10	AV524167	AV524167
32	192.6	29.5	675	13	B1182893	B1182893
33	189.6	29.0	451	9	AU227890	AU227890
34	183.4	28.1	507	9	AT777460	AT777460
35	183	28.0	671	13	B1959716	B1959716
36	178.6	27.4	628	13	BM269335	BM269335
37	176.4	27.0	308	12	BE804809	BE804809
38	172.6	26.4	701	12	BG043358	BG043358
39	172	26.3	564	13	B1160655	B1160655
40	167.8	25.7	590	13	BM099343	BM099343
41	167.6	25.7	600	13	BM341015	BM341015
42	167	25.6	704	14	BQ624976	BQ624976
43	166.4	25.5	501	10	AM221667	AM221667
44	161.6	24.7	425	10	AV556381	AV556381
45	158.6	24.3	581	13	B1950438	B1950438

#### ALIGNMENTS

RESULT 1  
AM004542  
LOCUS 701932172 A. thaliana, mixed source Arabidopsis thaliana cDNA clone  
DEFINITION 701932172. mRNA sequence.  
ACCESSION AM004542  
VERSION AM004542.1 GI:5851571  
KEYWORDS EST.

#### ORGANISM

thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

#### REFERENCE

1 (bases 1 to 529)  
Chen, J., Montoya, M., Chan, E., Mooney, M., Carreon, B., Gilliland, D., Wang, X., Hillman, J., Guebler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Noblrig, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.  
Arabidopsis thaliana Gene Expression Microarray  
Unpublished (1999)

#### TITLE

#### JOURNAL

#### COMMENT

Contact: David Smoller, Ph.D.  
Genome Systems, Inc., a wholly owned subsidiary of Incyte  
Pharmaceuticals, Inc.  
4633 World Parkway Circle, St. Louis, MO 63134, USA  
Tel: 877-577-2733  
Fax: 314-427-3324  
Email: service@genomesystems.com.

FEATURES  
source Location/Qualifiers  
1. .529  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/clone="701932172"  
/clone\_lib="A. thaliana, mixed source"  
/note="This sequence was obtained from a clone generated with a PCR product of the target gene."  
BASE COUNT 136 a 114 c 118 g 160 t 1 others  
ORIGIN

Query Match 76.8%; Score 501.4; DB 10; Length 529;  
Best Local Similarity 98.9%; Pred. No. 1.5e-146;  
Matches 526; Conservative 0; Mismatches 2; Indels 4; Gaps 2;

QY 101 CAAGTGAAGTGGCTCAATTAGTTCAGTAATCTGCGATCTTCAGATGGG 160  
DB 1 CAAGTGAAGTGGCTCAATTAGTTCAGTAATCTGCGATCTTCAGATGGG 60  
QY 161 CTGAAGTCMAAGCCGCTGCTCAAAATCATCTCAATTCGTTGAACCTCTCTAC 220  
DB 61 CTGAAGTCMAAGCCGCTGCTCAAAATCATCTCAATTCGTTGAACCTCTCTAC 120  
QY 221 CTCAGAAATGACTCTCTATACAGATGAATGCTGACCTGGAACAAGATTGGT 280  
DB 121 CTCAGAAATGACTCTCTATACAGATGAATGCTGACCTGGAACAAGATTGGT 180  
QY 281 ATCCGTTCTTATATACGAGCTCAGACGCTGGGCTGATGTATGATCATTCCTCTTGT 340  
DB 181 ATCCGTTCTTATATACGAGCTCAGACGCTGGGCTGATGTATGATCATTCCTCTTGT 240  
QY 341 CTGCTAACACATTAGCCAAAGATTGGTGGTATGTATGATCATTCATGACATGTATAG 400  
DB 241 CTGCTAACACATTAGCCAAAGATTGGTGGTATGTATGATCATTCATGACATGTATAG 300  
QY 401 TAAGACATGGATTAATACCAACCGTTGTTGTCACCGGCGATGAACACTTGTATGT 460  
DB 301 TAAGACATGGATTAATACCAACCGTTGTTGTCACCGGCGATGAACACTTGTATGT 360  
QY 461 GGAACAATCCCTTCAAGAAAGGACCTGCTGCTGATATGAACTTGAATACCTCTAA 520  
DB 361 GGAACAATCCCTTCAAGAAAGGACCTGCTGCTGATATGAACTTGAATACCTCTAA 417  
QY 521 TTCTCCCATCAAGAAGAACTGGCTGTGAGACTACGCTATGAGCGCAAT-GGCTGAG 579  
DB 418 TTCTCCCATCAAGAAGAACTGGCTGTGAGACTACGCTATGAGCGCAATGGCTGAG 477  
QY 580 CCTTCTGATTTATTCACCTGTTAGACTGTTCTGGGAGTCAACAGCTGTA 631  
DB 478 CCTTCTGATTTATTCACCTGTTAGACTGTTCTGGGAGTCAACAGCTGTA 529

RESULT 2  
A1994093 536 bp mRNA linear EST 08-SEP-1999  
LOCUS 701498662 A. thaliana, Ohio State clone set Arabidopsis thaliana  
DEFINITION cDNA clone 701498662, mRNA sequence.  
ACCESSION A1994093  
VERSION A1994093.1 GI:5840998  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 536)  
REFERENCE  
AUTHORS Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guebler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutou, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argente, C., Shah, S., Nobrigha, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.  
TITLE Arabidopsis thaliana Gene Expression Microarray

JOURNAL Unpublished (1999)  
COMMENT Contact: David Smoller, Ph.D.  
Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.  
4633 World Parkway Circle, St. Louis, MO 63134, USA  
Tel: 877-577-2733  
Fax: 314-427-3324  
Email: service@genomesystems.com.

FEATURES  
source Location/Qualifiers  
1. .536  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/clone="701498662"  
/clone\_lib="A. thaliana, Ohio State clone set"  
/note="cDNA library was made from selected clones from the Arabidopsis thaliana Ohio State clone set."  
BASE COUNT 137 a 118 c 109 g 172 t  
ORIGIN

Query Match 52.9%; Score 345.6; DB 9; Length 536;  
Best Local Similarity 84.9%; Pred. No. 1.9e-97;  
Matches 387; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 197 ATTGCTGATTAACCTCTCTACCTCAGATGCTCTATACAGATGAATGAAT 256  
DB 3 ATTTCTCGATTAACCTCTCTCCACAGAGTGAATGCTGATGATGAATGAAT 62  
QY 257 GGTCTGCTGGAACAAGATTGGATCCGCTTCTGATATGAGCTCAGACGCTGGGCTG 316  
DB 63 GGTCTGCTGGAACAAGATTGGATCCGCTTCTGATATGAGCTCAGACGCTGGGCTG 122  
QY 317 ATGTATGATCATTTGCTCTCTTGTCTGCTACACAAATTAACCAAGATTGCTGGTAT 376  
DB 123 ATGTATGATCATTTGCTCTCTTGTCTGCTACACAAATTAACCAAGATTGCTGGGCT 182  
QY 377 GGTATATCTATTTGATCATGTATAGTACATGGATGATTAACCAACCGTTGTTG 436  
DB 183 GGTATATCTTCTGATCATTTATGAGCTGTTGGACTATACCAACCAACCGTTGTTG 242  
QY 437 CACCGCGATGAACACTTTGATGTGGAACAATCCCTTCAAGAGGACCTGCTTTCG 496  
DB 243 CTCACACTGATGAATCTTGTATGTGGAACAATCCCTTCAAGAGGACCTGCTTTCG 302  
QY 497 TTGATGAACCTTGAATCACCTTAATTCCTCCATCAAGAAGAACTGGCTGTGAGACT 556  
DB 303 TTGATGAACCTTGAATCACCTTAATTCCTCCATCAAGAAGAACTGGCTGTGAGACT 362  
QY 557 ACGGTAATGGCGCAATGGCTGAGCTTCTGTGATTTATTCACCTGTTAGACTGTTGG 616  
DB 363 ACGGTAATGGCGCAATGGCTGAGCTTCTGTGATTTATTCACCTGTTAGACTGTTGG 422  
QY 617 AGTCACAGCTGTAACAAGAGATGAGAACCACTT 652  
DB 423 AGTCACAGCTGTAACAAGAGATGAGAACCACTT 458

RESULT 3  
B1204061 765 bp mRNA linear EST 11-JUL-2001  
LOCUS B1204061  
DEFINITION EST522101 cTOS Lycopersicon esculentum cDNA clone cTOS3N14 5' end,  
mRNA sequence.  
ACCESSION B1204061  
VERSION B1204061.1 GI:14681785  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 765)  
REFERENCE  
AUTHORS van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R., Roming, C. and Tanksley, S.

Generation of ESTs from Tomato Suspension Cultures  
Unpublished (2001)  
Contact: CUGI  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
location/qualifiers

BASE COUNT	210 a	133 c	345 2:	DB 13:	Length 765:
ORIGIN					

Query Match	32.5%	Score	220	
Best Local Similarity	76.2%	Pred. NO.	3.2e-97	
Matches 425; Conservative		Mismatches	133; Indels	0; Gaps
				0

[illegible]

## RESULT 4

SOURCE ORGANISM	potato.	Solanum tuberosum	Viridiplantae; eukaryota;	Embryophyta; Streptophyta; Viridicordales; core eudicots;	Tracheophyta;

REFERENCE  
AUTHORS

1 (bases 1 to 712) Sun, H., Cho, J., Chieming, S.,  
van der Hoeven, R., Bezzerides, J., Taksley, S. and Baker, B.  
Duell, C. R., Ronning, C.

**JOURNAL COMMENT**  
Unpublished (2003)  
Contact: Cathy Ronning  
The Institute for Genomic Research  
The Institute for Genomic Research

FEATURES	Seq primer: M13F-R:	Location/Qualifiers
1..712		"tuberosum"

```

/db_xref="taxon:4113
/clone="CSTS13J14"
/clone_id="CSTS"

```

```
/loc_host="SOLR"  
/lab_host="Vector"; site_1: EcCorr; site_2:  
/note="Bluescript SK(-); Site-1: EcCorr; site-2:  
          sizes of sprouting eyes (2mm to 15mm) in  
          incubated at 26C in
```

Accession	Year	Location	Altitude	Soil	Plant	Part	Condition	Remarks
203	1953	203 f						dark for 2 1/2" immediately upon freezing in liquid nitrogen tubers.

	ORIGIN	Length /L/;
Query Match	Score 335.6; DB 12;	
	Pred. No. 3.3e-94;	
	Indels 75.1%; Gaps 0;	
	Tidels 130.	

[illegible][illegible][illegible][illegible][illegible]

Dd  
403 CAGCAATCTTTCACGAAACGC

Gy  
425 CGTTGTTGTTCACCGGGCATGACACTTGATGTGAACAATCTTTCACGAAACGC

[illegible]

LOCUS DEFINITION	BE055140	823 bp	mRNA	linear	EST 07-MAR-2001
LOCUS	BE055140				
DEFINITION	BE055140	823 bp	mRNA	linear	EST 07-MAR-2001
	GA_Ea0033A05f	Gossypium	arborescens	7-10 dpa fiber library	Gossypium
	arborescens	CDNA clone	GA_Ea0033A05f	mRNA sequence.	

ACCESSION	BE055140
VERSION	BE055140.2
KEYWORDS	GI:13246149
SOURCE	EST. <i>Gossypium arboreum</i> .

ORGANISM	REFERENCE
<i>Gossypium arboreum</i>	1 (bases 1 to 823)
Eukaryotes: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Eukaryotes: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots: Rosidae: eurosids II: Malvales: Malvaceae: <i>Gossypium</i> .	

**AUTHORS** Ming, R.A., Frisch, D., Yu, Y., Mein, D., Rambo, T., Simmons, J., Henry  
D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
**TITLE** An integrated analysis of the genetics, development, and evolution  
of the cotton fiber  
**JOURNAL** Unpublished (2000)  
**COMMENT** On Jun 8, 2000 this sequence version replaced gi:8382197.

Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: [twing@clemson.edu](mailto:twing@clemson.edu)  
Seq primer: TAAATCGACATCACTATAGGG  
High quality sequence stop: 637.

FEATURES	Location/Qualifiers
source	1. .823

```

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0033A06f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/notes="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

```

ORIGIN			
Query Match	50.0%	Score 326.6;	D8 10; Length 823;
Best Local Similarity	73.5%;	Pred. No. 2.5e-11;	
Matches 430; Conservative	0;	Mismatches 154;	Indels 1; Gaps

44 AGAGATGCAATATGCAAGTGCATACAGTACACAGGAAGCCTCGTATCTTACTAGCTGCAA 103

Db  
20 ATAGGAGATGCTTAAAGTCATCTTACCCCAAGAAAACCCCGGTTTACTCGCTGCCA 79

104 GTGGAAGTGTGCGCTTCAATTAACTTCAGTAATCTCTGCCATTGTTCACAGATGGCGT 163

Db 80 GTGGAGTGTAGCTCCCATAGTTCGGAACTCTGCCATTGTTCCTCGATGGGAG 139

164 AAGTCAAGCCGTCGTTCAAAATCATCTCTCAATTTTCGTGATAAACCTTCTCTAACCTC 223

140 AGTAAAGCAGTTCGCCAGCAAGCTTCTTGCAATTTCATGACATAGCATTCACCTTCCCTA 199

QY	224	AGAAATGACACACCTCATTACACATGAAGATGATGGTACGTAACCAAGATTTGGTGATC	283
Db	200	AGGATCTAAAGCTTTACACTGATGAGAGGAATGGTCTAGTTGGGGGAAATTAAGGTAGCA	259
QY	284	CCGTTCTTCATATCGAGCTCAGACGCGTGGGGCTGATGTTATGATCACTTGCCTTTGCTCG	343
Db	260	GTGTCTTCACATGATGGCTGCTGATGGGCTGATATTAATGTCATATGCCCAATTTGAC	319
QY	344	CTAACACATTAAGCCCAAGATTGCTGGGTATATGATAATCTAATGACATGTAATAGTAA	403
Db	320	CAAAACACCTTGGCAAGATTGCTGAGGATTAATGACAAATTTGTTAACTGTGTGTAAC	379
QY	404	GAGCATGGGATTAATGACAACCGTTGTTTGGTTGTCACCGGGCGATGAAACATTTGATGTGA	463
Db	380	GAGCATGGGACATACAGCAAGCCAAATGTTTGTGGACACAGTATGAACAATTTCAATGTGA	439
QY	464	ACAATTCCTTTCACAGACGGCACCTTGCTTGGCTTGATGAATCACTTGGAAATCACCTTAATTC	523
Db	440	GCACACCTTTTCACAGAAAAGCATTCATGACAAATTAATGAGAGCTTGATTTCTCTCATCC	499
QY	524	CTTCCCATCAGAGAATCTGGCGCTGTGGAGACTAACGGTAATGGGCCAATGGCTGAGCCCT	583
Db	500	CCCCTGTCTCCAA - AGACTAAGCTTTGTGGGACATTAATGGGAACGGCCAAATGGCAGAACCTT	558
QY	584	CTCTGATTTAATCCACTGTTAGACTGTTTGGGAGTCAACCACTC	628
Db	559	CTCTAATCACTGAGACTGTAAAGATTATTCCTTGGAGTCAACGACCTC	603

RESULT 6	745 bp	mRNA	linear	EST 18-MAY-2001
AW218763				
LOCUS				
DEFINITION	AW218763			
	EST301243	tomato root during/after fruit set, Cornell University		
		EST301243 contig also aligns to a thaliana		

LYCOPERSICON ESCULENTUM CDM  
 HALS homology, mRNA sequence.  
 ACCESSION AM218763  
 VERSION AM218763.1 GI:6529637  
 KEYWORDS EST.  
 SOURCE tomato.  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Solanum

asteridae; eusteleis 1; solanates; solanaceae; solanum, lycopersicon.

REFERENCE	I (bases 1 to 43)	
AUTHORS	van der Hoeven, R.S.,	Garvin, D., Matern, A.L., Holt, I.E., Liang, F.,
	Waters, E.,	Graves, M.D., Bourne, C.I., Ahn, S., Bonning, C.M.

upton, J., Hansen, I., Clavett, M.B., Bowman, C.D., and S. J. Nordin, C.R. Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.

Generation of ESs from human foetal tissue  
Unpublished (1999)

COMMENT contact: CUGI  
Clemson University Genomics Institute  
at 2000 Redwood Rd.

Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: [htara@clemson.edu/orders/index.htm](mailto:htara@clemson.edu/orders/index.htm)

email: <http://www.genome.clemson.edu/olufsen/>  
5 prime sequence.

```

FEATURES
source
LOCATION/Qualifiers
1. .745
Location="Ruggero, John"

```

```

/organism="Lycopersicon esculentum"
/cultivar="TA496"
/lab="ucsf-tavon/0001"

```

```

/du_xier= laxou:4001
/clone="CLEXIEI5"
/clone_lib="+omato root during/after fruit set [Cornell]

```

/cione-lib-  
 comaco 1001 uiling/alber 1142 sec/ connect  
 University"  
 414964 tunc- "root"

```

/ribsue_type= root
/dev_stage="plants during and after fruit-set"
/notes="vector: nblinescript sk(-). Site 1: Froot1: Site 2:

```

```

/note - vector: printscript tms " string: books; over_21
Xhol; supplier: Tanksley; Tissue supplied by Dave Garvin
/ucba - abc 17thage NY 14850" #

```

BASE COUNT	196 a	155 c	170 g	224 t
CRISTIN	(USDA PMS, Litchfield, N. H. 1900).			

ORIGIN  
 Query Match 49 3% Score 322.2 DB 10 Length 745

QUEEN MARCH 1957

and XhoI was ligated to modified lambda FLU-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBlueScript vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further details.

```

FEATURES
source      Location/Qualifiers
1. .557    "Arabidopsis thaliana"
```

```
1. .55/  
/organism="Arabidopsis thaliana"  
/db_xref="taxon:3702"
```

```

/db_xref="taxon:3702"
/clone="RAFL15-30-H14"
/clone_lib="RAFL15"

```

```
/clone_lib="RAFL15"  
/tissue_type="mixture of silique and flower"  
/test="nuc10R"
```

```

/tissue_type="muscle"
/lab_host="DH10B"
/note="Site_1: BamHI; Site_2: SalI"
161 + 3 others

```

BASE COUNT	145 a	121 c	127 g	161 t	3 others
145 a	121 c	127 g	161 t	3 others	

BASE COUNT	145 d	121 c	127 c
ORIGIN	46.18	Score 320.8;	DB 9; Length 55

Query Match	Score	DB	Length
82.68;	320.8;	9;	55
82.68;	1.3e-89;		
82.68;	80;		

Query match			
Best Local Similarity	82.6%;	Pred. No. 1.3e-05;	
Matches	389;	Mismatches	80;
	Conservative		Indels

Matches 389; Conservative 87

53 ATATGCACTGGATACAGTACACAGGAGCCCTGATCTTACTAGCTGCAAGTG

[illegible]

Db  
88 ACATGAGTGAATACCAACCAGGAGGCCCTGGTACTACCCACCGA

QY 113 TGGCTCAATTAACTCAGTAATCTCGCCATTGTTTCTCAGAATGGGCAGAA

Db 148 TCGTGCTATCAAAATTCGGCAATCTCTGCCATTGCTTACCGAATGGCAGAAAC

Db  
148 TCGTCGCTCAAAATCATTCTGCAATTTCGTGATAAACCTTCTCTACCTCAG

DY  
173 CCGTCGCTCAAAATCATTCTGCAATTTCGTGATAAACCTTCTCTACCTCAG

[illegible]

	CCGTCGTACGAATCATCTTACATTCTCCAGAAATC
Db	208 CCGTCGTACGAATCATCTTACATTCTCCAGAAATC
	CTCTCTATACAGATGAAGTGAATGGTCTACGTCGAACAAGATTGGTGATCCC

QY  
233 CTCTATACAGATGAAGTGGTCCTACTCGAACACAGATTGTTGTAATCC

Db 268 CTCGTACTGATGAAGATGATGGTCTAGCTGACAAGATCGGTGATCCCT

DB 208 CTTTCTGTCCT  
QY 293 ATATGAGCTCAGACCGCTGGCTGATGTATGATCATTCCTTTGTCGCT

DY 293 AATAGCACTGCGTGGGTGTTTGCTCCT  
| | | | | | | | | | | | | | | |  
Db 328 ACATCGAGCTTAGACGTGGGCTGATGTTTAGTCATTGCTCTTGGCTCT

Dd 328 ACATCGAGCTTAGACGTGGGCGAAGCAATTTC

Ov 353 TAGCCNAGATTGCTGGTGGTTATGTGATAATCTATTGACATGTATAGTAAAGT

Q7 353 TAGCCAGATTGCTGGTGGTTATGTGAATACIAI ISACHA.....  
||| ||||||||||||||| ||||||||| ||| || | |||

Db 388 TAGGACAGATTGCTGGTGGCTTTGTGATAATCTTCTGACTTGCATTAACG

413 ATTATAGCAACCGTGTGTTTGTCACGGCAGTAGACAC-TTYGATGTGGA  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

413  
2Y  
448  
Db

ACTATACCAACCACTGTTGTGCTCCAGCTATGAATCTTTGATGTGGA

D<sub>b</sub> 448 ACTATACCAAAACCAGCTGTTTGGATTCCACCTAAT  
D<sub>v</sub> 472 TTTCACAGAAGGCACCTTGCTGTGATGAACTTGGAATCACCCTAAT

QY 472 TTTCACGAGAACGGCACCTTGTCTCTGCATCGAATC  
||||| ||| | | | | | | | | | | | | | | |  
508 TTTCACCTTGAA-GGNATTNTTTGTNTCTTGATGAACCTGGCAATCACA

Db 508 TTCACTCGA-GGNATNTTTTGTGNTCTTGATGAACTGGAGAACGC

## RESULT 8

	RESULT 8	
BQ273330		
BQ273330		
566 bp	mRNA	linear
	cDNA clone	SOYBEAN

LOCUS	DEFINITION	566 bp	mRNA	GENE
BQ273330	sao23h08.y1 Gm-cl054 Glycine max cDNA clone			SOYBEAN
	1054-5243 5' similar to TR:Q9SWE5			Q9SWE5 HAL3

**DEFINITION** *sao23h08.y1* *Gm cl054.5* similar to *TR:Q9SWE5 Q9SWE5 HAL3* *Gm-cl054-5343 5'* similar to *TR:Q9SWE5 Q9SWE5 HAL3* sequence.

sequence.  
BQ273330  
BQ273330.1 GI:20498400

ACCESSION	BQ273330	GI:20498400
VERSION	BQ273330.1	
KEYWORDS	EST.	

KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max

**SOURCE OF ORGANISM**

Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta  
Eukaryota; Viridiplantae; eudicotyledons; core  
eudicotyledons; Magnoliophyta; Magnoliopsida  
eu-magnoliopsids; Fabales; Fabaceae; Papiliono-

Eurycorymbaceae; Magnoliophyta; eudicotyledons, core  
Spermatophyta; Magnoliophyta; eudicotyledons, core  
Rosidae; eurosids I; Fabales; Fabaceae; Papiliono-

REFERENCE  
1 (bases 1 to 566)  
Glycine.  
Rosidae; eurosios 1, 1000000.  
Erpellding, J., C

**REFERENCE**

1 (bases 1 to 566)

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Crolla, B., Marra, M., Hillier, L., Kucaba, T., ... et al.

AUTHORS  
Snoekwaal, N., Hillier, L., Kucaba, I.,  
'A', Bolla, B., Marra, M., Theising, B.,  
Wyllie, T., Underwood, K., Steptoe, M.

Wyllie, T., Underwood, N., Deeprose, J.





Db	291	TATCTTACTAGCTGCAAGTGAAGTGGGCTTCATTTAAGTTCACTAATCTCTGCCATTG	232
Qy	147	TTTCTCAGAAATGGGCTGAAGTCAAAAGCCGTCCTTCAAAATATCTCTCAATTTTCGTGA	206
Db	231	TTTTCTCAAAATGGGCTGAAGTCAAAAGCCGTCGCTTCAAAATATCTCTCAATTTTCGTGA	172
Qy	207	TAAACCTTCTACTCGAAGATGTAGCTCTATTCACAGTGAAGTGAAGTGGTCTAGCTG	266
Db	171	TAAACCTTCTACTCGAAGATGTAGCTCTATTCACAGTGAAGTGAAGTGGTCTAGCTG	112
Qy	267	GAACAGATGTGGTGAATCCCGCTTCATATGAGAGCTCAGACGCTGGAGCTATGTTATGAT	326
Db	111	GAACAGATGTGGTGAATCCCGCTTCATATGAGAGCTCAGACGCTGGAGTGAAGTATGAT	52
Qy	327	CATTGCTCTCTTGTCTGCTTAACACATTAGCCAAG	360
Db	51	CATGGCTGATCTGTGTAGCGCATTAAGCCAAG	18
RESULT	10		
LOCUS	R30079	452 bp	mRNA
DEFINITION	12684 lambda-PRL2 Arabidopsis thaliana cDNA clone 164P177, mRNA		
ACCESSION	R30079		
VERSION	R30079.1	GI:936772	
KEYWORDS	EST.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eustoids II: Brassicales: Brassicaceae: Arabidopsis. 1 (bases 1 to 452)		
AUTHORS	Newman,T., deBurlin,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Rezel,E., and Somerville,C.		
TITLE	Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones Plant Physiol. 106, 1241-1255 (1994)		
JOURNAL MEDLINE	95148729		
COMMENT	On Apr 14, 1993 this sequence version replaced gi:785394. Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan State University MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI Tel: 517-353-0854 Fax: 517-353-9168 Email: 22313@cnsi.msu.edu Seq primer: T7 dye primer.		
FEATURES	Location/Qualifiers		
source	1..452		
	/organism="Arabidopsis thaliana"		
	/strain="var Columbia"		
	/db_xref="taxon:3702"		
	/clone="164P177"		
	/clone="164P177"		
	/note="Vector: lambda zip-Lox; Site_1: Sal; Site_2: Not; lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark - rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dT primed cDNA."		
BASE COUNT	100 a 100 c 99 g 135 t 18 others		
ORIGIN			
Query Match	47.6%	Score 311:	DB 14: Length 452:
Best Local Similarity	82.2%	Pred. No. 1,3e-86:	
Matches 347/	Conservative 0:	Mismatches 75:	Indels 0: Gaps 0:

OY	197	ATTTCGTTGAATCAACTTCTTACCTCAGAAATGACGTCTCATATGCATGAGAATGAT	256
Db	3	ATTTCCTCGATTAACCTCTCTCTCCACAAGANGTGACTCTMTATCTGATGAAGATGAAT	62
OY	257	GGTTCAGCTGGAACAAGATTGGTGATGCCGTTTCCTTCATATGAGAGCTCAGACGTGGGCTG	316
Db	63	GGNCTTAGCTGGAACAAAGATCGGTGATCCTTNCTTCNNACATGAGAGCTTAGAGCTTGGGCTG	122
OY	317	ATGTTATGATCATCTTCTCTCTTTGTCTGCTAACAATTAAGCCAAGATTGCTGGGTAT	376
Db	123	ATGTTTWTAGTCAATGCTCTCTTGTCTGTCAACACCCTTAVGGGAAGATTCCNGTGGGCTTT	182
OY	377	GTGATATATCTATGTCATGATGATAGTAAGAAGATGGAGATTATAGCAAACCGTTGTTGTTG	436
Db	183	GTNTATATCTCTCTGACTTGCATTAATACGACGCTTGGAGCATATACCAACCACTGTTTGTGG	242
OY	437	CACCGCGCATGAACAACCTTTGATGTGGAAACAATCCTTTTCACAGAACGACCTTGCTTTCG	496
Db	243	CTCCAGCATATGAATPACTTNNATGTGGAAACAATCCTTTTCACAGAAAGCATCTTTNGTCTC	302
OY	497	TTCATGAACCTTGGATTCACCTTAATTCCTCCCATCAACAAGAACTGGCCTGTGAGACT	556
Db	303	TTCATGAACCTGGGGGYCACACTTATTCCTCTCATCAAGAAAGNCTTCCTGTTGGGAGCT	362
OY	557	ACGGTAATGGGCGCAATGGCTGAGACCTTCTGATTTATTTCCAGTGTGAGCTGTCTGG	616
Db	363	ACGGTAATGGAGCTATGGCTGAGACCTTCTTATTTATTTCCAGCTGTGAGACTTTTNTGG	422
OY	617	AG	618
Db	423	NG	424
RESULT	11		
LOCUS	AY105313	1165 bp	mRNA linear HTC 25-MAY-2002
DEFINITION	Zea mays PCO119881 mRNA sequence.		
ACCESSION	AY105313		
VERSION	AY105313.1 GI:21208391		
KEYWORDS	HTC.		
SOURCE	Zea mays.		
ORGANISM	Zea mays.		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 1165) Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V. Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes Unpublished (2002)		
JOURNAL	2 (bases 1 to 1165)		
REFERENCE	Coe,E.C.		
AUTHORS	Direct Submission		
TITLE	Submitted (25-APR-2002) Maize Mapping Project; University of Missouri, Columbia, MO 65211, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1..1165		
source	/organism="Zea mays" /db_xref="MaizEDB:636974" /db_xref="taxon:4577" /clone="PCO119881" /clone_idb="Maize Mapping Project/Dupont Consensus Library" /note-"this sequence is part of a project of EST assemblies resulting from the application of public attribs to seed Dupont configs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACS in conjunction with the Maize Mapping Project"		
BASE COUNT	291 a 275 c 307 g 292 t		
ORIGIN			







GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2002, 19:19:30 ; Search time 98 Seconds  
(without alignments)  
2566.126 Million cell updates/sec

Title: US-09-868-300-7  
Perfect score: 1 gaattcgacagagctcctt.....caaagatggaaccagttg 653  
Sequence: 1 gaattcgacagagctcctt.....caaagatggaaccagttg 653

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131.6	20.2	383	10	US-09-878-574-5053 Sequence 5053, App
2	121.2	18.6	903	9	US-09-764-904-34 Sequence 34, App1
3	121.2	18.6	903	10	US-09-764-860-293 Sequence 293, App
4	60.2	9.2	4503	10	US-09-070-927A-149 Sequence 149, App
5	57.6	8.8	281	10	US-09-878-574-6517 Sequence 6517, App
6	52	8.0	11842	9	US-09-764-904-126 Sequence 126, App
7	52	8.0	11842	10	US-09-764-860-1187 Sequence 1187, App
8	48.2	7.4	1218	10	US-09-974-300-984 Sequence 984, App
9	44	6.7	787	9	US-09-895-913A-65 Sequence 65, App1
10	34.4	5.3	394	10	US-09-878-919-37 Sequence 37, App1
11	33.2	5.1	387	10	US-09-878-574-1174 Sequence 1714, App
12	33.2	5.1	15567	12	US-10-047-676A-3 Sequence 3, App1
13	33	5.1	1067	10	US-09-974-300-5426 Sequence 5426, App
14	33	5.1	4527	10	US-09-901-940-3 Sequence 3, App1
15	32.4	5.0	398	10	US-09-983-965-5883 Sequence 5883, App
16	32.2	4.9	277	9	US-10-046-935-2165 Sequence 2165, App
17	32.2	4.9	692	10	US-09-764-869-1990 Sequence 1990, App
18	31.6	4.8	568	10	US-09-764-869-1991 Sequence 1991, App
19	31.6	4.8	578	10	US-09-764-869-384 Sequence 384, App

20	31.6	4.8	2000	9	US-09-938-842A-2966 Sequence 2966, App
21	31.2	4.8	1738	10	US-09-925-300-294 Sequence 294, App
22	30.8	4.7	574	10	US-09-864-761-12057 Sequence 12057, App
23	30.8	4.7	1188	10	US-09-867-550-2027 Sequence 2027, App
24	30.6	4.7	377	10	US-09-895-828-237 Sequence 237, App
25	30.6	4.7	2000	9	US-09-938-842A-3829 Sequence 3829, App
26	30.4	4.7	557	9	US-10-046-935-359 Sequence 359, App
27	30.4	4.7	5194	9	US-10-002-389-1 Sequence 1, App1
28	30.2	4.6	374	10	US-09-878-574-3522 Sequence 3522, App
29	30.2	4.6	418	10	US-09-960-352-14362 Sequence 14362, App
30	30.2	4.6	1268	10	US-09-070-927A-557 Sequence 557, App
31	30.2	4.6	1272	10	US-09-808-483-7 Sequence 7, App1
32	30.2	4.6	30013	10	US-09-764-877-3297 Sequence 3297, App
33	30	4.6	675	10	US-09-974-300-6755 Sequence 6755, App
34	30	4.6	32038	10	US-09-764-878-292 Sequence 292, App
35	29.8	4.6	396	10	US-09-864-761-4628 Sequence 4628, App
36	29.8	4.6	396	10	US-09-864-761-21370 Sequence 21370, App
37	29.8	4.6	475	10	US-09-867-701-2393 Sequence 2393, App
38	29.8	4.6	3293	8	US-08-910-386A-17 Sequence 17, App1
39	29.8	4.6	6101	10	US-09-972-714-1 Sequence 1, App1
40	29.6	4.5	325	9	US-09-920-455-6 Sequence 6, App1
41	29.6	4.5	360	10	US-09-895-828-231 Sequence 231, App
42	29.4	4.5	346	10	US-09-070-927A-752 Sequence 752, App
43	29.4	4.5	427	9	US-09-920-455-153 Sequence 153, App
44	29.4	4.5	612	10	US-09-974-300-5525 Sequence 5525, App
45	29.4	4.5	1432	10	US-09-822-830A-394 Sequence 394, App

#### ALIGNMENTS

RESULT 1  
US-09-878-574-5053  
Sequence 5053, Application US/09878574  
Patient No. US20020110548A1  
GENERAL INFORMATION:  
APPLICANT: Byrum, Joseph R.  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Thompson, Michael D.  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(1540)B  
CURRENT APPLICATION NUMBER: US/09/878,574  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 09/333,535  
PRIOR FILING DATE: 1999-06-14  
NUMBER OF SEQ ID NOS: 15775  
SEQ ID NO 5053  
LENGTH: 383  
TYPE: DNA  
ORGANISM: Glycine max  
OTHER INFORMATION: Clone ID: LIB3028-011-Q1-B1-D1  
US-09-878-574-5053

Query Match 20.2%; Score 131.6; DB 10; Length 383;  
Best Local Similarity 65.1%; Pred. No. 3.5e-33;  
Matches 194; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY	40	GAGAGAAGATGATATGAGAGTGAAGTACAGTACAGGAAGCTCTATCTACTAGCT	99
DB	86	GAATGAGAAAACCAACCAAGTGTGATGCTGCCACAGGACCCGGGATTCATCTCT	145
QY	100	GCAAGTGAAGTGGTGGCTTCATTAAGTTCAGTATCTGCGCCATTTCTCTCGAATGG	159
DB	146	CCAATTGGAATGTGGGGCTGTAATAATTTGCAATTTTGTCTACTGTTTCCGATGG	205
QY	160	GCTGAAGTCAAAAGCGTGGCTTCAAAATCATCTCTCAATTTCTGTGTAACCTTCTCTA	219
DB	206	GCAATGTAAATGAGATTTCCACAAAGTCAATCTTGCATTTCAATTAAGAGCAGCAATG	265
QY	220	CCTCAGATGTGATCTCTATACAGATGAAGATGAATGGTCTAGCTGGAACAGATTGGT	279
DB	266	CCCAAGATGTAATTCATACAGATGACATGAATGAATGGTCTACTTGAAGAAATTAAGT	325



```

? FILING DATE: 1997-05-06
? APPLICATION NUMBER: 60/066,009
? FILING DATE: 1997-11-14
? ATTORNEY/AGENT INFORMATION:
? NAME: Kenley K. Hoover
? REGISTRATION NUMBER: 40,302
? REFERENCE/DOCKET NUMBER: PB369
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (301) 309-8504
? TELEFAX: (301) 309-8512
? INFORMATION FOR SEQ ID NO: 149:
?     SEQUENCE CHARACTERISTICS:
?         LENGTH: 4503 base pairs
?         TYPE: nucleic acid
?         STRANDEDNESS: double
?         TOPOLOGY: linear
? ?
?-S-09-070-927A-149      SEQUENCE DESCRIPTION: SEQ ID NO: 149:
```

Query Match	9.2%	Score	60.2	DB	10	Length	4503
Best Local Similarity	50.5%	Pred.	No.2.9e-09				
Matches 146; Conservative	0;	Mismatches	143;	Indels	0;	Gaps	0

QY	287	TTCTTATATTCGAGCTCAGACGCTGGGCTGATGATATGATCATTCCTCTTGTCTGCTA	346
Db	2325	TTAATCAATATGAGTAGCTAAAAAAGCGGATTTATTTTTTATAGTACACTCGCTCGCGCA	2266
QY	347	ACCAATTAGCCAAGATTGCTGCTGCTTATGTGATATATCTATTGCATGTATAGTAGAG	406
Db	2265	ATTCAAATTGGGAATATATGACCCATGGGATTTGCCAGACACTATTATTAACAACAGTTCCTTGG	2206
QY	407	CATGGATTAATAGCAACCGTTGTTGGTGACGGCGCATGAAACCTTTGATGTGAGACA	466
Db	2205	CACCTTATATCTGAACACCAACCAAAATTTATTCGCCCGCGGTGATACGTATATGTACCAA	2146
QY	467	ATCCTTTGACAGAGGCACTTGCTTCTGCTGATGAACTTGGAATCAACCTTAATTCCTC	526
Db	2145	ACCCCAATTGTCACAGCAAAATATTGGCAATTTTAAAGAAGTCGGTATACAGAAATATATTC	2086
QY	527	CCATCAAGAAAGAACTGGGCTGTGGAGACTACGGTATATATGGCGCAATGGC	575
Db	2085	CACGTGAACCTCTATTAGCTTGTGGCGATTATAGGTGCGGCGCACCTTGC	2037

```

RESULT 5
US-09-878-574-6517
: Sequence 6517, Application US/09878574
: Patent No. US20020110548A1
: GENERAL INFORMATION:
: APPLICANT: Byrum, Joseph R.
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Thompson, Michael D.
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(15401)B
: CURRENT APPLICATION NUMBER: US/09/878,574
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 09/333,535
: PRIOR FILING DATE: 1999-06-14
: NUMBER OF SEQ ID NOS: 15775
: SEQ ID NO 6517
: LENGTH: 281
: TYPE: DNA
: ORGANISM: Glycine max
: OTHER INFORMATION: Clone ID: 701098531H1
: US-09-878-574-6517

```

[illegible]

Db 178 GACATATGGCTGTGGATGCTGCCCCCAAGGAAACCCCCGATTCCTACTTCTCGAGTGGGAG 238  
Oy 111 TGTGGCTCAATTAAGTTCAGTAATCTCTGGCAATGTTTCTCAG 154  
238 TGTTCCTGCTCTCAAAATTTGCAAAATCTTTTGCACTGTTTCTCTG 281

```

RESULT 6
US-09-764-904-126
: Sequence 126, Application US/09764904
: Patent No. US20020173454A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: Pal22
: CURRENT APPLICATION NUMBER: US/09/764,904
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ. ID NOS: 137
: SOFTWARE: PatentIn Ver. 2.0
: SEQ. ID NO 126
: LENGTH: 11842
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-764-904-126

```

Query Match	8.0%:	Score 52:	DB 9:	Length 11842:
Best Local Similarity	56.4%:	Pred. No. 2.3e-06:		
Matches 97:	Conservative 0:	Mismatches 75:	Indels 0:	Gaps 0:
QY 390	GACATGTATACGTAAAGCATGCATGGATATATAGAAACCGTGTGTTTGTTCACCGCGCATGAA	449		
Db 6819	GACCTGCGTCAATGCGGGGCGCTGGAGCCGACGACGACGCCCTGCTCTTCTGCGCGGCATGAA	6878		
QY 450	CACTTTGTGTGGAACAATCTTTTCACAGAAAGGCACCTGTGCTTGTGTGATGAACCTGG	509		
Db 6879	CACCGCCATGTGGGAGGACCCCGATCACAGCCAGCAGGATAGACCAGCTCAAGGCCCTTTGG	6938		
QY 510	AATCACCTATTCTTCCTCCCATCAAGAAACAAATGGCCCTGTGAGACTACGCT	561		
Db 6939	CTATGTCCAGATCCCTCTGTGTGGCCAAACAACCTGTGTGGCGAGATGGAAGT	6990		

```

RESULT 7
US-09-764-860-1187
; Sequence 1187 Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ. ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1187
; LENGTH: 11842
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-860-1187

```

	Query Match	Similarity	Score	DB	Length	11842
	Best Local	Similarity	56.48	Pred.	2.3e-06	
	Matches	97	Conservative	0	Mismatches	75
				Indels	0	Gaps
						0
QY	390	GACATGTATAGTAAAGACATGGGATTTATAGCAAAACCGTTGTTTGGCACCAGCGCATGAA	449			
Db	6819	GACCTGGCTACTGGGGGCTCTGGACCGAGAGAGCCCTCGCTCTTTCGCCGGCCATATGAA	6878			
QY	450	CACTTTCATGTGGAACATCCTTTCACAGAACGGCACCTTCTCTTGGCTTGTATGAACTTGG	509			
Bb	6879	CACCGCATGTGGGAGACACCGCATACAGCCACAGGATAGACCAGCTCAAGGCCCTTGG	6938			







SEQ ID NO 3  
LENGTH: 4527  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: This degenerate nucleotide sequence encodes the  
OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.  
NAME/KEY: misc\_feature  
LOCATION: (1)...(4527)  
OTHER INFORMATION: n = A,T,C or G  
US-09-901-940-3

Query Match 5.1%; Score 33; DB 10; Length 4527;  
Best Local Similarity 23.6%; Pred. No. 2.2;  
Matches 73; Conservative 61; Mismatches 175; Indels 0; Gaps 0;

OY 11 CGAGCTCCTCCTCGGCTGTAACAAGATGAGACAGATGATGAGATGAGTACAG 70  
DB 212 CMTGCMNMGNTTNNMNSNCARGATNAARGARGARYTNACNARGACNAGCTNC 271  
OY 71 TAACAAGGAGGCTGCTATCTAGCTGCAAGTGAAGTGGCTTCAATTAGTTCA 130  
DB 272 ARGCNTTMSNTTNTNGARYNGAYGCAATTAATTAATTAATTAATTAATTAAT 331  
OY 131 GTAATCTGCGCATGTTTTCAGATGGCTGAAGTCAAGCCGCTGCTCAAAATCAT 190  
DB 332 GYWSNGCMNSNGTNTTYYTNGARAAYCCNCAVGTNCARWSNGTNCNATHGARWSNC 391  
OY 191 CTCTCAATTTCTGTTAAACCTTCTTACCTCAGAAATGTGACTCTCTATACAGATGAAG 250  
DB 392 ARGATTTTTCGNGCMNTHAARYTNCARCNGARYTMSNACNATHMSNGARAYGGNA 451  
OY 251 ATGAATGCTAGTGAACAAGATTTGATCCGTTCTCATATGAGCTCAGACGCT 310  
DB 452 ARGATATATATYTNMGNATHGARWSNACNGTNCNATHATHTGYWSNGARTYWSNGARY 511  
OY 311 GGGCTGATG 319  
DB 512 TNGAYCARG 520

RESULT 15  
US-09-983-965-5883/c  
Sequence 5883, Application US/09983965  
Patent No. US20020137160A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengding  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 37-21(10297)C  
CURRENT APPLICATION NUMBER: US/09/983, 965  
CURRENT FILING DATE: 2001-10-26  
PRIOR APPLICATION NUMBER: US 09/465,231  
PRIOR FILING DATE: 1999-12-15  
PRIOR APPLICATION NUMBER: US 60/113,678  
PRIOR FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 5912  
SEQ ID NO 5883  
LENGTH: 398  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
OTHER INFORMATION: Clone ID: 63-LIB34-021-Q1-E1-H4  
US-09-983-965-5883

Query Match 5.0%; Score 32.4; DB 10; Length 398;

Best Local Similarity 51.4%; Pred. No. 0.89;  
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

OY 116 CTTCATTAGTTCAGTATCTGCCATTGTTCTCAGAAATGGCTGAAGTCAAGCCG 175

DB 229 CTTCAATTAAATTTCAAGTGTGATTTGATTTCTGTAATTTGCTGCTTTCCGTG 170  
OY 176 TCGCTCAAAATCATCTCTCAATTTGATTAACCTTCTCTACCTGAGATGAGTACTC 235  
DB 169 TCAATTTGAACCTTCTCAAGTTTGTGTTTCTTCCATCAGATGAGTTAGTCTT 110  
OY 236 TCTATACAGATGAGATGAATGCTCT 261  
DB 109 TAATTAAGAGAGACTCATTTTCT 84

Search completed: December 1, 2002, 21:57:41  
Job time : 123 secs



QY	339	GTGCGAACACATTTATTCACAAAGATTTGGTGGCTTATGTGATTAATCTATTTCACATGTAT	398
Db	9062	AACGCTACACATTTTGCACAAACTAGCTACAGSATTTGGGACACATGTAACAGTA	9122
QY	399	AGTAAGACATGGGATTTATAGCAAAACGGTGTTTGGTTCACGGCGATGAAACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACCAAGTATATTTCCAAACTAAATAGTCCCTGATGTAAATACAAAAT	9181
QY	459	GTGGACAATCTCTTCACACACAGGACCTGTCTGTGTGTAAGAACTGGTAATACCT	518
Db	9182	GTATGACCATTCAGTAATCTGGAATAACATTAGCACTACGGCTATATA -GCT	9240
QY	519	AATTCCTCCATCAAGAAACTGGCGTGTGAGACTACGGTAATGCGCAATGGCTGA	578
Db	9241	GATTGTCTCTAAGAAATCCCTACTAGCTGTGAGACCAAGACGAGAGCTTTAGCTGA	9300
QY	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

```

RESULT 2
US-09-134-001C-1948
: Sequence 1948, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1948
: LENGTH: 1203
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-1948

```

Query Match	8.48;	Score 55;	DB 4;	Length 1203;
Best Local Similarity	50.28;	Pred. No. 1.2e-08;		
Matches 167;	Conservative	0;	Mismatches 160;	Indels 6;
				Gaps 1

OY	272	AGATTGGGATCCCGTCTTCATATACGAGCTCACACCGCTGGCGGATGTTATGATCATTTG	331
Db	194	AAATTCCTGGAAGAGATTACATGATCATTCATTAGAGACATGGGCGAGATCGATTATAGTCG	253
OY	332	CTCCCTTTCTGCTCAACACATATACCAAGATTGCTGGGGGTATATGATTAATCTATTGA	392
Db	254	CGCCAGCAACCTGCTATATCTATCGCAAAATTAAAGTGTGGACATTGCTGATGATTTAATTA	313
OY	392	CATGTATAGTAAAGCACAGGAGATTATAGCAACCGTTGTTTGGACCGCGCATGACA	451
Db	314	CTTCTACTTACTTGC-----TACAAACACCAAAATTCGTTGGACCCGCAATGATG	367
OY	452	CTTTGATGTGCAACATCTTTTCACAGACGGCACCTTGCTGCTGATGAACTTGAA	511
Db	368	TAAATATATTAACAATCCACGTACTAAACATATATGAAAGTGCATATGCAACAGCGAT	427
OY	512	TCACCCATATTCCTCCCATCAAGAGAAACTGGCCTGTGAGACTACGTAATAGCGCAA	571
Db	428	ATTATTTATTTGAACCTGGTAGTGCGCTATTAGCATGTGGTTATGTAGCAAAAGCGCAA	487
OY	572	TGGCGAGCCCTTCTGATTTATTCCACTGTA	604
Db	488	TGGAAGAACCCATGCAAAATCTTACTGTATTA	520

### RESULT 3

```

US-09-134-001C-1999/c
; Sequence 1939, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1939
; LENGTH: 207
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1939

```

[illegible]

RESULT 4  
US-08-822-774-18  
; Sequence 18, Application US/08822774

1 APPLICANT: HOGREFE, Holly  
2  
3 TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
4 TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF-Proteins  
5  
6 TITLE OF INVENTION: and Methods for Purifying and Identifying Same  
7  
8 NUMBER OF SEQUENCES: 61  
9  
10 CORRESPONDENCE ADDRESS:  
11  
12 ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &  
13 ADDRESSEE: Lenahan, P.L.C.  
14  
15 STREET: 1200 G Street, N.W. Suite 700  
16  
17 CITY: Washington  
18  
19 STATE: D.C.  
20  
21 ZIP: 20005  
22  
23 COMPUTER READABLE FORM:  
24  
25 MEDIUM TYPE: Floppy disk  
26  
27 COMPUTER: IBM PC compatible  
28  
29 OPERATING SYSTEM: PC-DOS/MS-DOS  
30  
31 SOFTWARE: PatentIn Release #1.0, Version #1.25  
32  
33 CURRENT APPLICATION DATA:  
34  
35 APPLICATION NUMBER: US/08/822,774  
36  
37 FILING DATE: 21-MAR-1997  
38  
39 CLASSIFICATION: 436  
40  
41 ATTORNEY/AGENT INFORMATION:  
42  
43 NAME: KULIK, David J.  
44  
45 REGISTRATION NUMBER: 36,576  
46  
47 REFERENCE/DOCKET NUMBER: 1486/43163  
48  
49 TELECOMMUNICATION INFORMATION:  
50

US-08-822-774-18

Query Match	6.6%	Score 43.4	DB 4	Length 1209
Best Local Similarity	51.3%	Pred. No. 0.0001		
Matches 101; Conservative	0	Mismatches 96	Indels 0	Gaps 0

[illegible]

RESULT 5  
US-09-632-711-18  
; Sequence 18, Application US/09632711  
Date: 09-09-2009

GENERAL INFORMATION:

TITLE OF INVENTION: Polym

and Methods for Purifying and Identifying Same  
NUMBER OF SEQUENCES. 61

CORRESPONDENCE ADDRESS:

Lenahan, P.L.L.C.

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

APPLICATION NUMBER: US/09/632,711

CLASSIFICATION: <Unknown>

APPLICATION NUMBER: 08/822,774

ATTORNEY/AGENT INFORMATION:  
NAME: RUTH T. DAVIES

REGISTRATION NUMBER: 36,576  
REFERENCE/DOCKET NUMBER: 1406/423163

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-8800

TELEFAX: (202) 628-8

LENGTH: 1209 base pairs

```

:      type: nucleic acid
:      STRANDEDNESS: single
:      TOPOLOGY: unknown
:      MOLECULE TYPE: DNA (genomic)
:      HYPOTHETICAL: NO
:      ANTI-SENSE: NO
:      SEQUENCE DESCRIPTION: SEQ ID NO: 18
US-09-632-711-18

```

Query Match	6.6%	Score 43.4	DB 4	Length 1209
Best Local Similarity	51.3%	Pred. No. 0.0001		
Matches 101	Conservative 0	Mismatches 96	Indels 0	Gaps 0

[illegible]

RESULT 6  
US-09-632-703B-18  
; Sequence 18, Application US/09632703B

GENERAL INFORMATION:

TITLE OF INVENTION: Polym

Identifying Same  
NUMBER OF SEQUENCES, 61

;  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Elnora Henderson  
 Freshport, Connecticut  
 December 11, 1906

STREET: 1300 I S  
CTRY: Washington

STATE: D.C.  
ZIP: 20005-3315

```

; COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk

```

COMPUTER: <Unknown>  
OPERATING SYSTEM: I

```

: SOFTWARE: PatentIn Release #1.0,
:
: CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/632,703E  
FILING DATE: 24-Aug-2000

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/822,114  
FILING DATE: 21-MAR-1997

ATTORNEY/AGENT INFORMATION:  
NAME: BARKER, M. Paul

REGISTRATION NUMBER: 32,013  
REFERENCE/DOCKET NUMBER: 4121.0116-04

TELEPHONE: (650) 849-6613

TELEPHONE: (800) 845 0000  
; INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHANGES LISTED:  
LENGTH: 1209 base pairs

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527

ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-632-703B-18

Query Match 6.6%; Score 43.4; DB 4; Length 1209;  
Best Local Similarity 51.3%; Pred. No. 0.0001;  
Matches 101; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 313 GCTGATGTTATGATCATGCTTCCTTGTCTGCTAACATTAAGCAACATTTGCTGTGG 372  
DB 277 GCAGATTAAATTTGGTTGTCTGCTCCACGTCCACACAAATTAAGATTGCAATGTGA 336  
QY 373 TTATGTATATATCATATGACATGTATAGAGCATGGATTATAGCAACCGTTGTTT 432  
DB 337 ATAGATGATATCTCCAGTACATACAGTGTGACACAGCATTTCCCAATTCATTTATG 396  
QY 433 GTTGACCCGGGATGACACATTTGATGTGGAACATCTTTACAGAAAGGACCTTGTG 492  
DB 397 ATAGCCCGAGCATGACAGCAATGTACAGGATCCCATAGTAAAGGAGAACATTTGAA 456  
QY 493 TTGCTGATGAACCTTGG 509  
DB 457 AGTTAAAGAACCTTGG 473

## RESULT 7

US-09-632-702-18  
Sequence 18; Application US/09632702  
Patent No. 6444428

## GENERAL INFORMATION:

APPLICANT: HOGREFE, HOLLY  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
and Methods for Purifying and Identifying Same

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESSES:

ADDRESSEE: David J. Kulik, Evenson, McKown, Edwards &  
Lenahan, P.L.L.C.

STREET: 1200 G Street, N.W. Suite 700

CITY: Washington

STATE: D.C.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/632,702

FILING DATE: 04-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/822,774

FILING DATE: 21-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Kulik, David J.

REGISTRATION NUMBER: 36,576

REFERENCE/DOCKET NUMBER: 1486/43163

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-8800

TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 1209 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-632-702-18

Query Match 6.6%; Score 43.4; DB 4; Length 1209;  
Best Local Similarity 51.3%; Pred. No. 0.0001;  
Matches 101; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 313 GCTGATGTTATGATCATGCTTCCTTGTCTGCTAACATTAAGCAACATTTGCTGTGG 372  
DB 277 GCAGATTAAATTTGGTTGTCTGCTCCACGTCCACACAAATTAAGATTGCAATGTGA 336  
QY 373 TTATGTATATATCATATGACATGTATAGAGCATGGATTATAGCAACCGTTGTTT 432  
DB 337 ATAGATGATATCTCCAGTACATACAGTGTGACACAGCATTTCCCAATTCATTTATG 396  
QY 433 GTTGACCCGGGATGACACATTTGATGTGGAACATCTTTACAGAAAGGACCTTGTG 492  
DB 397 ATAGCCCGAGCATGACAGCAATGTACAGGATCCCATAGTAAAGGAGAACATTTGAA 456  
QY 493 TTGCTGATGAACCTTGG 509  
DB 457 AGTTAAAGAACCTTGG 473

## RESULT 8

US-09-221-017B-363  
Sequence 363; Application US/09221017B  
Patent No. 6444799

## GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

INFORMATION FOR SEQ ID NO: 363:

SEQUENCE CHARACTERISTICS:

LENGTH: 4086 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

```

: ANTI-SENSE: UNKNOWN
: ORIGINAL SOURCE:
: ORGANISM: PORPHYROMONAS GINGIVALIS
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1...4086
: US-09-221-017B-363

Query Match          6.6%; Score 43; DB 4; Length 4086;
Best Local Similarity 64.6%; Pred. NO. 0.00028;
Matches 64; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 292 CATATCGAGCTCAGACGCGTGGGCTGATCTTATGATCATTCGTCCTTTGTCTCTAACA 351
   ||| ||| ||||| ||| ||||| ||||| ||| ||||| ||| ||| |||
Db 2473 CATGTGATCTCAGCTCTCTGGCGGATGCTATGCTGATGACGCGCACACGCTCTACG 2532
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 352 TTAGCCAGATTGCTGCTGCTGGTTATGATATATATCATTTG 390
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2533 ATAGGCAGATGGCAGCATGTCATACCCACATATGTTG 2571

RESULT 9
US-08-392-625-16
: Sequence 16, Application US/08392625
: Patent No. 5837485
: GENERAL INFORMATION:
: APPLICANT: Entlan, Karl-Dieter
: APPLICANT: Gtz, Friedrich
: APPLICANT: Schnell, No. 5837485bert
: APPLICANT: Augustin, Johannes
: APPLICANT: Engelke, Gernar
: APPLICANT: Rosenstein, Ralf
: APPLICANT: Kleita, Corina
: APPLICANT: Klein, Cora
: APPLICANT: Wieland, Bernd
: APPLICANT: Wipke, Thomas
: APPLICANT: Jung, G nther
: APPLICANT: Kellner, Roland
: TITLE OF INVENTION: Biosynthetic Process For The Preparation
: TITLE OF INVENTION: Of Chemical Compounds
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1100 New York Avenue
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/392.625
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/876,791
: FILING DATE: 30-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: REFERENCE/DOCKET NUMBER: 0652.0980002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8700 base pairs
: type: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```





NAME: Mabury, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 222/005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2976 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
OS-08-714-918-65

Query Match	5.38;	Score 34.8;	DB 3;	length 2976;
Best Local Similarity	50.08;	Pred. No. 0.14;		
Matches 144;	Conservative 0;	Mismatches 137;	Indels 7;	Gaps 2

Qy	310	TGGGCTATCTTTGATGACATTCGCTCTTTGTGCGTAACACATTTGACCAAGTTCTGGT	369
Db	1755	TGGGCMATTCATCTCANTGTTCGACCTCGCAAGCGCAATACATTTGCCAAATTTGAGTGA	1698
Qy	370	GCGTTATGATTAATCTAATTTAGACATGTATAGTAAGACATGGGATTTATAGCAACCGTTG	429
Db	1695	GGTATTCGATATATTGTGGACATCAACGTGTGCTGACACAGAG-----ACACCGAAA	1642
Qy	430	TTTGTTCACCGCGGATGAAACACTTTGATGTGGGAACAATCCTTTACAGAAACGGACCTT	489
Db	1641	TTTTATTCGCTCTATGAAATGTGCAATATGTATGAAAATTAACGTACCCACACAAATAT	1582
Qy	490	GTCCTTGCTATCAACTTGGAAATCACCACTTAATTCCTCCCATTAAGAAGA-NACTGGCCTG	548
Db	1581	TAAATATTTTAAAGAAGATGGGATCATTTTATTCGAACCAAGGAAGGATTTCTTGATG	1522
Qy	549	TGGAGACTACGGTAATGGCGCAATGGCTAGCCTTCTCGATTTATTC	596
Db	1521	TGGTATAGTTGCTTAAGAGACGTATGGAAACACCGCTTCAAACTCGTTTC	1474

RESULT 13  
 US-09-265-315-65/C  
 Sequence 65 Application US/09265315  
 Patent No. 6187541  
 GENERAL INFORMATION:  
 APPLICANT: Benton, Bret  
 APPLICANT: lee, Yung J.  
 APPLICANT: Malouin, Francois  
 APPLICANT: Martin, Patrick K.  
 APPLICANT: Schmid, Molly B.  
 APPLICANT: Sun, Dongxu  
 TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
 TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
 TITLE OF INVENTION: TARGET GENES  
 NUMBER OF SEQUENCES: 111  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Lyon & Lyon  
 STREET: 633 West Fifth Street  
 STREET: Suite 4700  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 90071-2066  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
 MEDIUM TYPE: storage  
 COMPUTER: IBM compatible  
 OPERATING SYSTEM: IBM P.C. DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/265,315  
 FILING DATE: March 9, 1999  
 CLASSIFICATION: 435

```

P R I O R   A P P L I C A T I O N   D A T A :
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Waldburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2976 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-65

```

Query Match	5.38;	Score 34.8;	DB 4;	Length 2976;
Best Local Similarity	50.08;	Pred. No. 0.14;		
Matches 144;	Conservative 0;	Mismatches 137;	Indels 7;	Gaps 2

QY	310	TGGCGCATGTTTATGATCATTCGCTCCCTTGTCTGCTATACACATTTGACCAAGATTCGTGT	369
Db	1755	TGGSCMATGCATCATCATTTGTTGCACTCGCAGCGGCAATATCATTTGCCAAATTTGAGTGTGA	1696
QY	370	GCGTTATGTATATCTATTTGACATGTATATAGTAAGACATGGGATTTATAGCAACCGTTG	429
Db	1695	GGATTTGCTATATTTATTTGGATCATCAACGTTGCTGTCAACAGAG-----ACACGGAA	1642
QY	430	TTTGTTCACCGCGGATGAACACCTTTGATGTGGGAACAATTCCTTTCACGAAGCGGCACTT	489
Db	1641	TTTATTTGCGCTGCTATGAAATGTGCATATGATATGAAATATTAACGTAACCGACCAAAATAT	1587
QY	490	GTCCTTGCTGATGAACCTTGGAAATCACCTTAATTCCTCCCATCAAGAGA- AACTGCGCTG	548
Db	1581	TAAATATTTTAAAGAAAGATGGTATCATTTTATTCGAACCGAAGGAGGATTTTGTACATG	1522
QY	549	TGGAAGCATACGGTAATGCGCCAAATGGCTGAGCCTTCTCGATTTATTC	596
Db	1521	TGGTATGTTGCTAAAGGACGTATGGGAAGAACCGCTTCAAACTGTTCT	1474

RESULT 14  
 US-09-265-315-65/c  
 : Sequence 65, Application US/09265315  
 : Patent No. 6187541  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Benton, Bret  
 :  
 : APPLICANT: Lee, Ying J.  
 : APPLICANT: Malouin, Francois  
 : APPLICANT: Martin, Patrick K.  
 : APPLICANT: Schmid, Molly B.  
 : APPLICANT: Sun, Dongxu  
 :  
 : TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
 :  
 : TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
 :  
 : TITLE OF INVENTION: TARGET GENES  
 :  
 : NUMBER OF SEQUENCES: 111  
 :  
 : CORRESPONDENCE ADDRESS:  
 :  
 : ADDRESSEE: Lyon & Lyon  
 :  
 : STREET: 633 West Fifth Street  
 :  
 : STREET: Suite 4700  
 :  
 : CITY: Los Angeles  
 :  
 : STATE: California  
 :  
 : COUNTRY: U.S.A.  
 :  
 : ZIP: 90071-2066  
 :  
 : COMPUTER READABLE FORM:  
 :  
 : MEDIUM TYPE: 3.5" Diskette, 1.44 MB

MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/265,315  
FILING DATE: March 9, 1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/714,918  
FILING DATE: September 13, 1996  
APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Walburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 240/247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2976 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-265-315-65

Query Match 5.3%; Score 34.8; DB 4; Length 2976;  
Best Local Similarity 50.0%; Pred. No. 0.14;  
Matches 144; Conservative 0; Mismatches 137; Indels 7; Gaps 2;

OY 310 TGGCGTATGTTATGATGATCTTCCTTGTCTGCTAACACATTAGCCAGATTGCTGT 369  
DB 1755 TGGCGMATGCAATCATCTTGTGACATCAACGCGCAATACAAATTCAGTGT 1696  
OY 370 GGGTATGATGATATCTTATGACATGTATGATAGAGACATGGATTTATAGCAACCGTTG 429  
DB 1695 GGTATGCTGATGATGTTGCTACATCAACGTTGCTAGCAACGAG-----ACACCGAAA 1642  
OY 430 TTTGTTGACCGGGGATGACACTTTGATGTGACAACTCTTCACAGAACGCGACTT 489  
DB 1641 TTTATTTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1582  
OY 490 GTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 548  
DB 1581 TAAATTTTAAAGAGATGGGTATCATTTATGACACAGAGGAGGATTTCTAGCATG 1522  
OY 549 TGGAGACTAGCGTAATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 596  
DB 1521 TGGTATGTTGCTTAAAGAGATGATGAGAGAACCGCTTCAAAATCGTTTC 1474

RESULT 15  
US-09-266-417-65/c  
Sequence 65. Application US/09266417  
Patent No. 6228588  
GENERAL INFORMATION:  
APPLICANT: Benton, Bret  
APPLICANT: Lee, Ving J.  
APPLICANT: Malouin, Francois  
APPLICANT: Martin, Patrick K.  
APPLICANT: Schmid, Molly B.  
APPLICANT: Sun, Dongxu  
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
TITLE OF INVENTION: TARGET GENES  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon

STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/266,417  
FILING DATE: March 9, 1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/714,918  
FILING DATE: September 13, 1996  
APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Walburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 240/248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2976 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-266-417-65

Query Match 5.3%; Score 34.8; DB 4; Length 2976;  
Best Local Similarity 50.0%; Pred. No. 0.14;  
Matches 144; Conservative 0; Mismatches 137; Indels 7; Gaps 2;

OY 310 TGGCGTATGTTATGATGATCTTCCTTGTCTGCTAACACATTAGCCAGATTGCTGT 369  
DB 1755 TGGCGMATGCAATCATCTTGTGACATCAACGCGCAATACAAATTCAGTGT 1696  
OY 370 GGGTATGATGATATCTTATGACATGTATGATAGAGACATGGATTTATAGCAACCGTTG 429  
DB 1695 GGTATGCTGATGATGTTGCTACATCAACGTTGCTAGCAACGAG-----ACACCGAAA 1642  
OY 430 TTTGTTGACCGGGGATGACACTTTGATGTGACAACTCTTCACAGAACGCGACTT 489  
DB 1641 TTTATTTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1582  
OY 490 GTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 548  
DB 1581 TAAATTTTAAAGAGATGGGTATCATTTATGACACAGAGGAGGATTTCTAGCATG 1522  
OY 549 TGGAGACTAGCGTAATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 596  
DB 1521 TGGTATGTTGCTTAAAGAGATGATGAGAGAACCGCTTCAAAATCGTTTC 1474

Search completed: December 1, 2002, 19:19:27  
Job time : 105 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2002, 17:42:20 ; Search time 302 Seconds  
(without alignments)  
4869.387 Million cell updates/sec

Title: US-09-868-300-7

Perfect score: 653  
Sequence: 1 gaattgcgcagcagctcctt.....caaaagatggaaccagttg 653

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/emb1/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/emb1/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/emb1/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/emb1/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/emb1/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/emb1/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/emb1/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/emb1/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/emb1/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/emb1/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/emb1/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/emb1/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/emb1/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/emb1/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/emb1/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/emb1/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/emb1/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/emb1/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/emb1/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/emb1/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/emb1/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/emb1/NA2001A.DAT:\*  
23: /SID52/gcgdata/geneseq/emb1/NA2001B.DAT:\*  
24: /SID52/gcgdata/geneseq/emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653	100.0	653	21	AAA51411
2	443.2	67.9	973	21	ABAC35405
3	151.6	23.2	3348	24	ABAC40569
4	128	19.6	500	21	ABAC93745
5	121.2	18.6	903	22	AA530612
6	121.2	18.6	903	22	AA528151
7	120.6	18.5	784	22	AAH05519
8	120.6	18.5	1208	21	AA334608
9	120.6	18.5	1699	22	ABA08875

C	10	120.6	18.5	1699	22	AA160412	Human polynucleoti
	11	120.6	18.5	1748	22	AA158626	Human polynucleoti
	12	120.6	18.5	2211	22	AAH15923	Human cDNA sequenc
	13	81.2	12.4	543	24	ABN69125	Streptococcus poly
	14	76.6	11.7	540	24	ABN69124	Streptococcus poly
	15	63.6	9.7	9909	19	AAV52145	Streptococcus pneu
	16	61.4	9.4	349980	21	AAAF21611	Neisseria meningit
	17	61.4	9.4	837096	21	AAAB1489	N. meningitidis pa
	18	60.2	9.2	4503	20	AAAX13086	Enterococcus faeca
	19	59.8	9.2	20844	21	AAAB1460	N. meningitidis pa
	20	59.8	9.0	676	21	AAA01916	Human colon cancer
	21	55	8.4	1200	22	AAH53591	S. epidermidis ope
	22	55	8.4	1203	24	ABN92485	Staphylococcus epi
	23	55	8.4	2937	22	AAH54809	S. epidermidis gen
	24	55	8.4	3150	22	AAH54529	S. epidermidis gen
	25	52	8.0	11842	22	AA530674	DNA encoding novel
	26	52	8.0	11842	22	AA528753	Genomic sequence #
	27	50.8	7.8	5692	23	AB194546	Drosophila melanog
	28	49.8	7.6	92407	22	AAAF28349	Genomic fragment #
	29	48.2	7.4	1218	24	ABK73693	Bacillus licheniflo
	30	44.8	6.9	5137	18	AAV74528	Staphylococcus epi
	31	44	6.7	207	24	ABN92476	Staphylococcus epi
	32	44	6.7	787	19	AAAX14137	H. pylori GHP0 319
	33	43.4	6.6	1209	19	AAV53859	Polymerase enhanci
	34	41.6	6.4	300	20	AAZ12612	Human gene express
	35	41.4	6.3	349980	22	AAH41224	Pyrococcus abyssi
	36	40.4	6.2	1260	22	AAH66739	C. glutamicum codin
	37	40.4	6.2	1389	22	AAAF72065	Corynebacterium gl
	38	40.4	6.2	349980	22	AAH68529	C. glutamicum codin
	39	38.4	5.9	4445	24	ABA01096	B. subtilis
	40	37.6	5.8	8700	14	AAQ42541	B. subtilis
	41	37.6	5.8	8700	14	AAQ42541	B. subtilis
	42	34.8	5.3	2976	21	AA525496	S. epidermidis plas
	43	34.8	5.3	2976	21	AA525496	Essential Staphylo
	44	34.8	5.3	2976	22	AA508065	Staphylococcus aur
	45	34.6	5.3	1231	23	AA564225	Staphylococcus aur

#### ALIGNMENTS

RESULT 1	
AA51411	AA51411 standard; cDNA: 653 BP.
ID	AA51411
XX	AA51411:
AC	26-SEP-2000 (first entry)
XX	
DE	A. thaliana Vb89 (HAL3) cDNA.
XX	
XX	Cyclin-dependent protein kinase; CDK: CDC2a; Vb89: HAL3;
KW	cell cycle; interacting protein; environmental stress; growth regulator;
KW	herbicide; nematode resistance; plant breeding; ss.
XX	
OS	Arabidopsis thaliana.
XX	
FH	key
FT	1..651
FT	location/Qualifiers
FT	/*tag= a
FT	/product= CDC2b-interacting-protein
FT	/partial
XX	
PN	WO200036124-A2.
XX	
PD	22-JUN-2000.
XX	
PE	17-DEC-1999; 99WO-EPI0084.
XX	
PR	17-DEC-1998; 98EP-0124062.
XX	
PA	(CROP-) CROPDESIGN NV.
XX	

PI De Veylder L, Boudolf WKCK, Torres Acosta JA, Inze D:  
XX WPI: 2000-431601/37.  
DR P-PSDB: AAY96816.  
XX  
PT Nucleic acids encoding plant cell cycle interacting proteins, useful  
for regulating plant growth and in recombinant DNA protocols  
XX  
PS Claim 1: Page 125-126; 152pp: English.  
CC The Vb89 clone encodes the Arabidopsis thaliana HAL3 homologue, a  
CC halotolerant gene isolated in *Saccharomyces cerevisiae*. The Vb89 clone  
CC interacts with A. thaliana CDC2b (a cyclin-dependent protein kinase  
CC (CDK)), but not with CDC2a in the two-hybrid system.  
CC CDC2a and CDC2b are the only CDK genes to have been characterized in  
CC detail in Arabidopsis thaliana. They were used as bait in a two-hybrid  
CC screening assay with a cDNA library of a plant cell suspension as prey.  
CC The plant cell cycle interacting proteins identified were designated  
CC LbV15, Phb80-like protein, Vb33, Vb89, VbDAHP and VbHSF. The nucleic  
CC acids, vectors comprising them, the proteins they express, antibodies  
CC that bind to them and or inhibitors of their protein expression and/or  
CC activity may be used for modulating the cell cycle in an animal or plant,  
CC plant cell division and/or growth, for influencing the activity of cell  
CC cycle proteins in a plant or animal cell, as positive or negative  
CC regulators of cell proliferation, for modifying the growth inhibitive  
CC caused by environmental stress conditions (e.g. to improve growth of  
CC plants in normal or suboptimal nutrient conditions, especially  
CC phosphorus), for use in a screening method for inhibitors or activators  
CC of cell cycle protein, as growth regulators, herbicides and/or for  
CC inducing nematode resistance in plants. The DNA sequences and their  
CC regulatory sequences may be used as markers in plant or animal cell and  
CC tissue cultures or as a marker in marker-assisted plant breeding. The  
CC regulatory sequences may also be used for the expression of heterologous  
CC DNA sequences during a stage of the cell cycle.  
XX  
SQ Sequence 653 BP, 180 A; 135 C; 151 G; 187 T; 0 other:  
  
Query Match 100.0%; Score 653; DB 21; Length 653;  
Best Local Similarity 100.0%; Pred. No. 8.5e-199;  
Matches 653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 481 CGCACCTGTGCTCTGCTGATGAGACTGTGAATCACCCTTAATTCCTCCATCAAGAAGAAA 540  
|||||  
Db 481 CGGCACCTGTGCTCTGCTGATGAGACTGTGAATCACCCTTAATTCCTCCATCAAGAAGAAA 540  
Qy 541 CTGGCCTGTGAGACTAGCTAATGGCGCAATGCTGAGCCCTCCATTTATTCACCT 600  
|||||  
Db 541 CTGGCCTGTGAGACTAGCTAATGGCGCAATGCTGAGCCCTCCATTTATTCACCT 600  
Qy 601 GTTAACCTGTTCTGCGACTCACAAGCTCGTAACAAAGAGATGGAACAGTTG 653  
|||||  
Db 601 GTTAACCTGTTCTGCGACTCACAAGCTCGTAACAAAGAGATGGAACAGTTG 653  
  
RESULT 2  
AAC35405  
ID AAC35405 standard; DNA; 973 BP.  
XX  
AC AAC35405;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SPQ ID NO: 10075.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
PD  
XX 06-SEP-2000.  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139819.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144326.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144633.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154039.  
PR 22-SEP-1999; 99US-0154779.  
PR 23-SEP-1999; 99US-0155139.  
PR 24-SEP-1999; 99US-0155486.  
PR 26-SEP-1999; 99US-0155659.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

```
QY 53 ATATGAGTGGATACAGTAAAGAGCCGTATCTTACTACTGCAAGTGAAGTG 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 ACATGGAAGTGAATRCCACCGAGAGAGCCCTCGTACTACTGCTGCAAGTGAAGCG 190
QY 113 TGCGTTCAATTAAAGTCAATCTCTGCAATGTTGTTCTCGAATGGGCTGAAGTCAAG 172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 TCGGTGATCAAAATTCGGCAATCTCTGCAATGTTGTTCTCGAATGGGCTGAAGTCAAG 250
QY 173 CCGTGGCTTCAAAATTCATCTCTGCAATGTTGTTGATTAACCTTCTTACTACTGAGATGGA 232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 CCGTGGTACCAAAATTCATCTCTGCAATGTTGTTGATTAACCTTCTTACTACTGAGATGGA 310
QY 233 CTCTCTATACAGATGAAGATGAATGTTGTTGATTAACCTTCTTACTACTGAGATGGA 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311 CTCTGTATACAGATGAAGATGAATGTTGTTGATTAACCTTCTTACTACTGAGATGGA 370
QY 293 ATATGAGCTCAGACGCTGGGCTGATGTTATGATCAATGTTCTCTGCTTAACACAT 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 ACATGAGCTTACAGCTGGGCTGATGTTATGATCAATGTTCTCTGCTTAACACAT 430
QY 353 TAGCCAGATTTGCTGGGCTGATGTTATGATCAATGTTCTCTGCTTAACAGCATGGG 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 431 TAGCCAGATTTGCTGGGCTGATGTTATGATCAATGTTCTCTGCTTAACAGCATGGG 490
QY 413 ATATAGCAAAACCGTTGTTGTTGTCACCGCGCATGACACTTTGATGGAACAATCCCT 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 491 ACTATACCAAAACCGTTGTTGTTGTCACCGCATGATGATGTTGATGGAACAATCCCT 550
QY 473 TCACAGAAAGCGACCTTCTTCTGCTTGAATGAATGGAATCAACCTTAATCTCTCCATCA 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 TCACAGAAAGCGACCTTCTTCTGCTTGAATGAATGGAATCAACCTTAATCTCTCCATCA 610
QY 533 AGAAGAACTGGCCGTGGAGACTAGCTAATGAGCGCAATGGCTGAGCTTCTGATTT 592
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 AGAAGAGCTTGGCTGTGGAGACTAGCTAATGAGCTAATGAGCGCTTCTTATCT 670
QY 593 ATTCACCTGTTAGACTGTTGTTGAGAGTCAACAAGCTGTAACAAAGATGAGAACAGTT 652
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 ATTCACCTGTTAGACTGTTGTTGAGAGTCTACAGCTCATCAGCAAAACGGTGAATAGTT 730

RESULT 3
ABLA0569
ID 3 ABLA0569 standard; DNA; 3348 BP.
XX
AC ABLA0569;
XX
DT 17-JUN-2002 (first entry)
XX
DE Canarypox virus (CapV) non-essential region nucleotide fragment #1.
XX
KW Canarypox virus; CapV; recombinant; vaccine; gene; ds.
XX
OS Canarypox virus.
XX
PN JP2002045184-A.
XX
PD 12-FEB-2002.
XX
PF 01-AUG-2000; 2000JP-0233097.
XX
PR 01-AUG-2000; 2000JP-0233097.
XX
PA (JAPG ) NIPPON ZEON KK.
XX
WPI: 2002-299189/34.
XX
PT A DNA region nonessential to the growth of a virus and a recombinant
  canaripox virus by using it -
XX
PS Claim 1; Page 8-9; 14pp; Japanese.
XX
```

CC The invention provides a DNA region nonessential to the growth of a  
CC canarypox virus (CapV). A recombinant CapV, in which an exotic gene is  
CC inserted to the above DNA region nonessential to the growth of a CapV,  
CC can be used in an expression vector and a vaccine. The present sequence  
CC represents a nucleotide fragment of the DNA region non-essential to the  
CC growth of the canarypox virus.

XX Sequence 3348 BP; 1278 A; 442 C; 521 G; 1107 T; 0 other;

Query Match 23.2%; Score 151.6; DB 24; Length 3348;  
Best Local Similarity 62.8%; Pred. No. 9.5e-38;  
Matches 235; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
QY 233 CTCTCTATACAGATGAAGATGAATGTTGTTGATTAACCTTCTTACTACTGAGATGGA 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 289 CTATTTATCCGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 348
QY 293 ATATGAGCTCAGACGCTGGGCTGATGTTATGATCAATGTTCTCTGCTTAACACAT 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 349 ATATGAGCTTAAAGATGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATG 408
QY 353 TAGCCAGATTTGCTGGGCTGATGTTATGATCAATGTTCTCTGCTTAACAGCATGGG 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 409 TAGCCAGATTTGCTGGGCTGATGTTATGATCAATGTTCTCTGCTTAACAGCATGGG 468
QY 413 ATATAGCAAAACCGTTGTTGTTGTCACCGCGCATGACACTTTGATGGAACAATCCCT 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 469 ATATTAATAAACCTTACTTCTGCTTGAATGAATGGAATGGAATGGAATGGAATGGAATGGA 528
QY 473 TCACAGAAAGCGACCTTCTTCTGCTTGAATGAATGGAATGGAATCAACCTTAATCTCTCCATCA 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 529 TTACTAGACACCATTAATAAAGCTTGAATGAGCTTGAATGAGCTTGAATGAGCTTGAATG 588
QY 533 AGAAGAACTGGCCGTGGAGACTAGCTAATGAGCGCAATGGCTGAGCTTCTGATTT 592
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 569 AGAAGAACTGACCTGCTGGGCTGATGTTGAGAAAGGTGCATGACAGTACTGATATTG 648
QY 593 ATTCACCTGTTAGA 606
    |||||
Db 649 CACGTGTTGTCAGA 662

RESULT 4
AAC93745/C
ID 1 AAC93745 standard; cDNA; 500 BP.
XX
AC AAC93745;
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:240.
XX
KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
  flea infestation; vaccine; antiparasitic; therapeutic target;
  diagnosis; detection; ss.
XX
OS Clenoccephalides felis.
XX
PN WO200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09437.
XX
PR 09-APR-1999; 99US-0128704.
XX
PA (HESK-) HESKA CORP.
XX
WPI:
XX
PT Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
  Flea Malpighian tubule and head and nerve cord tissue derived nucleic
  acid
XX
```

PT acids useful for the prevention, diagnosis and treatment of flea  
infestations -

Claim 26: Page 325; 964pp; English.

CC The invention relates to novel cat flea (*Ctenocephalides felis*) nucleic  
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue  
CC or head and nerve cord (HNC) tissue. The invention also relates to the  
CC encoded proteins. The invention additionally encompasses expression  
CC constructs, recombinant viruses and recombinant cells comprising the  
CC nucleic acids of the invention, recombinant production of the proteins,  
CC antibodies against the proteins, a method of identifying inhibitors of  
CC the proteins, and compositions comprising the inhibitors for  
CC administration to an animal. The nucleic acids, and the proteins they  
CC encode may be used in the prevention, treatment and diagnosis of diseases  
CC associated with flea infestations. For example, the nucleic acids may be  
CC used to produce an HMT or HNC protein according to standard recombinant  
CC DNA methodology by inserting the nucleic acids into a host cell and  
CC culturing the cell to express the protein. The HMT and HNC nucleic acids  
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect  
CC and quantitate the presence of cat flea or other homologous nucleic acid  
CC sequences in samples. They may also be used to study the expression and  
CC function of the proteins and their role in metabolism. The HMT and HNC  
CC proteins may be used as antigens in the production of specific  
CC antibodies, and in assays to identify modulators (agonists and  
CC antagonists) of HMT and/or HNC protein expression and activity. The  
CC anti-HMT/HNC protein antibodies and antagonists may also be used to  
CC downregulate protein expression and activity. The antibodies may also be  
CC used as diagnostic agents for detecting the presence of flea polypeptides  
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The  
CC present sequence represents a cat flea HMT cDNA of the invention.

XX SQ Sequence 500 BP; 131 A; 91 C; 96 G; 181 T; 1 other;

Query Match 19.6%; Score 128; DB 21; Length 500;

Best Local Similarity 54.6%; Pred. No. 1.4e-30;

Matches 254; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY 182 CAAATGATCTCTCAATTTCTGTTGATTAACCTTCTACCTCAGATGTCGCTCTATA 241  
DB 466 CAGATATGCAAAACACTCTTTCGACCAATCAGAAATACCCCAACGTCAAATATATCA 407  
QY 242 CAGATGAAGATGATGCTGACGTAACCAAGATTGGTATCCGTTCTTATATCGAGC 301  
DB 406 CAGATGACCAATGAAATTTTGAACAAACGTTGGACCTGTTTACACATCGATT 347  
QY 302 TCAACGCTGGGCTGATGATATGATTCCTTCTGTTGCTAACAACATTAAGCAAGA 361  
DB 346 TAACATAATGGCGAGACATCATGTCATTGCACCGTTGGACGCCAACACTTTGGCTAAAA 287  
QY 362 TTGCTGGTGGATATGATTAATCTATTTGACATGTAATAGTAAGACATGGGATATAGCA 421  
DB 286 TATCAATATGGTTATGTAACAACTCTTAACCTGCACAGCTCGAGCCGGAACACATCCA 227  
QY 422 AACGCTGTTTGTTCGACCGGCGATGAACACTTTGATGTGGAACAATCTTTACAGAAC 481  
DB 226 AGCCCCCTCTGTTTTCGACCAATGAAATAGAGAAATGGGATACCCCAATACCCGCTC 167  
QY 482 GGCACCTTGCTCTGATGAATCTGGAATCACCTTAATCTCCCATCAAGAAAGAAC 541  
DB 166 CACACAGTAATAACTTTAAAGATGGGGGATATATANAAGTCCCTTGCACTTTCTAAGACTT 107  
QY 542 TGGCCTGTGAGACTACGTAATGCGCAATGGCTGAGCCCTTCTGATTTATTCACCTG 601  
DB 106 TGGTGTGCGGATAGTATGTAATGCTGATGCTGATGTAATATACATGTTGATGTG 47  
QY 602 TTGACATGTTCTGGGAGTCACAGCTGTAAACAAAGATGGA 646  
DB 46 TAAATGATGAGTTGAATAAAAAAAAAAAAAAAAAAAAAA 2

RESULT 5  
AAS30612

ID AAS30612 standard; cDNA; 903 BP.

XX AAS30612;

XX 21-NOV-2001 (first entry)

DE DNA encoding novel lung cancer antigen, Seq ID No 34.

KW Human; lung cancer; immunosuppressive; antiarthritic; antineumatic;  
KW antiproliferative; cytostatic; cardiac; vasotropic; cerebroprotective;  
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;  
KW ophthalmological; autoimmune disease; rheumatoid arthritis; neoplasm;  
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; skin aging;  
KW ocular disorder; wound healing; organ transplantation; ss.

XX Homo sapiens.

XX MO200155300-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01238.

XX 31-JAN-2000; 2000US-0179065.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-465565/50.

XX P-PSDB; AAU18952.

PT Isolated nucleic acid molecule encoding a lung cancer antigen is used  
in preventing, treating or ameliorating a medical condition -

XX Claim 1; SEQ ID No 34; 475pp; English.

CC The invention relates to novel isolated lung cancer antigen  
CC polypeptides (I) and polypeptides (II). (I) and (II) are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) are  
CC also used in diagnosing a pathological condition or susceptibility to a  
CC pathological condition, in particular, lung cancer. The antibodies to  
CC (II) can also be used in alleviating symptoms associated with the  
CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme  
CC linked immunosorbent assays (ELISA). Disorders which are diagnosed or  
CC treated include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
CC ocular disorders e.g. corneal infection. The polypeptides can also be  
CC used to aid wound healing and epithelial cell proliferation, to prevent  
CC skin aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis. AAS30580-AAS30685 represent novel human lung cancer antigen  
CC coding sequences. PCR primers and related sequences of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at: ftp.wipo.int/pub/published\_pat\_sequences.

XX SQ Sequence 903 BP; 194 A; 236 C; 271 G; 198 T; 4 other;

Query Match 18.6%; Score 121.2; DB 22; Length 903;

Best Local Similarity 59.0%; Pred. No. 2.8e-28;

Matches 207; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 228 TGTGACCTCTATACAGATGAATGATGCTGAGTAACAAGATTGGTATCCCGT 287  
DB 320 TGTGACCTCTATACAGCAGCGTGAATGGAGATGTGGAAGACCGCTGTGACCAAGT 379  
QY 288 TCTTCATATGAGCTCAGACGCTGGCTGATGATGATCATCTCTCTTGTCTGCTAA 347

DB 380 TCTGCACATTGACCTGCGAGGCGGACAGACCCCTGCTGAGTGCCTCTTGATGCCAA 439  
QY 348 CACATTAGCCAGATTCGGTGGGTTATGTATATCATTCATTCATGATAGTAGAGC 407  
DB 440 CACTCTGGGAGAGGTGGCAGTGGCATCTGTACACTTGCCTGATGGGGC 499  
QY 408 ATGGATTATAGCAACCCCTTGTGTTCACCGCGCATGACACTTGTATGTAACAA 467  
DB 500 CTGGAGCCGACAGAACCCCTGCTTCTTGCCGGCATGAAACCGCCATGTGGAGCA 559  
QY 468 TCCTTTCACAGAGGCGACCTTGTCTTGTGATGACTTGGATACCTTAATTCCTCC 527  
DB 560 CCCGATCACAGGCGCACAGTACCCAGCTCAAGGCTTGGTATGTGAGATCCCTG 619  
QY 528 CATCAGAGAAACCTGCGCTGAGACTAGGCTATGCGGCAATGGCTGA 578  
DB 620 TGTGGCCAGAAAGCTGTGTGCGAGATGAAGCTCTGGGGCCATGGCTGA 670

## RESULT 6

AAS28151

ID AAS28151 standard: cDNA; 903 BP.

XX AAS28151;

DT 07-NOV-2001 (first entry)

XX Novel cDNA encoding for human respiratory antigen #283.

XX Human; respiratory antigen; respiratory disorder; throat disorder;  
KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;  
KW anti allergic; anti asthmatic; anti inflammatory; olfactory;  
XX respiratory active; ss.

OS Homo sapiens.

XX WO200155448-A1.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01333.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.





CC In gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX  
SQ Sequence 784 BP; 166 A; 219 C; 229 G; 167 T; 3 other;

Query Match 18.5%; Score 120.6; DB 22; Length 784;  
Best Local Similarity 59.0%; Pred. No. 4.1e-28;

Matches 207; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 228 TGTGACCTCTATACAGTGAAGTGAATGCTGCTGAGCAACAGATTGGTATCCCGT 287  
DB 304 TGTACCCCTTACACGACGCTGATGAATGGAGATGTGGAAAGCCGCTTGACCCAGT 363  
QY 288 TCTTCAATTCAGCTCAGACGCTGAGTGTATGATCATCTGCTTGTCTGCTAA 347  
DB 364 TCTGACATTGACCTCGCGAGGTGGCAGACCTCCTGCTGGTCTCTCTTGATGCCAA 423  
QY 348 CACATTAACCAAGATTGCTGCTGCTTATGATATCATATGACATGATAGTAAGAC 407  
DB 424 CACTCTGGGGAAGTGGCAGTGCATCTGTCAAACTTGCTTACCTGCTGATCGGGC 483  
QY 408 ATGGATTATAGCAACCTGTTGTTGTGACACCGCGCATGAACTTGATGATGACAA 467  
DB 484 CTGGGACCGCAGCAAGCCCTGCTCTTGTGCGCCGCAATGAACCCGCTATGGAGCA 543  
QY 468 TCTTTACAGAGACGACCTTGTCTGCTGATGAACCTTGAAATCACCCTTATCTCC 527  
DB 544 CCCGATCAGACGCGCAGAGTACAGCTCAAGCCCTTGGCTATGTCGAGATCCCTCG 603  
QY 528 CATCAAGAAGAACTGGCTGTGAGACTACGTAATGCGCCATGCGTGA 578  
DB 604 TGTGCGCAAGAACTGTGTGCGAGATGAAGTCTCGGGGCCATGGCTGA 654

RESULT 8

AAZ34608 standard; cDNA: 1208 BP.

AAZ34608:

DT 15-FEB-2000 (first entry)

XX Human receptor molecule (REC) Incyte clone 2022379.

XX Receptor; REC: human; diagnosis; therapy; neoplastic disorder;

KW Immunological disorder; reproductive disorder; nervous disorder;

KW gastrointestinal disorder; smooth muscle disorder;

KW musculoskeletal disorder; ss.

OS Homo sapiens.

XX Key 71..1051 Location/Qualifiers

FT CDS /tag= a

PN MO957270-A2.

PD 11-NOV-1999.

PE 28-APR-1999; 99WO-US09191.

PR 01-MAY-1998; 98US-0071822.

XX (INCYTE) INCYTE PHARM INC.

XX Hillman JL, Bandman O, Tang YT, Yue H, Lal P, Corley NC;

PI Guegler KJ, Patterson C;  
XX  
XX WPI: 2000-052971/04.  
DR P-PSDB; AAY32199.

PT Novel human receptor molecules used in the diagnosis, treatment and  
PT prevention of neoplastic, immunological, reproductive gastrointestinal,  
PT nervous, smooth muscle and musculoskeletal disorders  
XX  
XX  
PS Claim 7; Page 86; 94pp; English.

CC This is the nucleotide sequence of Incyte cDNA clone 2022379  
CC encoding a novel human receptor molecule (REC, see AAY32199). The  
CC invention provides human RECs and polynucleotides which identify  
CC and encode REC, as well as vectors, host cells, antibodies,  
CC agonists and antagonists. Human RECs appear to play a role in  
CC neoplastic, immunological, reproductive gastrointestinal, nervous,  
CC smooth muscle and musculoskeletal disorders. The protein, antagonists  
CC and agonists, and compositions can be used to treat: a reproductive  
CC disorder, including but not limited to, prolactin production disorders,  
CC infertility including tubal disease, ovulatory defects, endometriosis,  
CC disruptions of the oestrous and menstrual cycles, polycystic ovary  
CC syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian  
CC tumours, uterine fibroids, autoimmune disorders, ectopic pregnancy,  
CC teratogenesis, breast cancer, fibrocystic breast disease,  
CC galactorrhoea, disruptions of spermatogenesis, abnormal sperm  
CC physiology, testis cancer, prostate cancer, benign prostatic  
CC hyperplasia, prostatitis, Peyronie's disease, male breast carcinoma  
CC and gynecomastia; gastrointestinal disorders including, but are not  
CC limited to, dysphagia, peptic oesophagitis, oesophageal spasm and  
CC stricture, oesophageal carcinoma, dyspepsia, indigestion, gastritis,  
CC gastric carcinoma, anorexia, nausea, emesis, gastroparesis, intestinal  
CC tract infection, peptic ulcer, colitis, Whipple's disease, Mallory-Weiss  
CC syndrome, irritable bowel syndrome, short bowel syndrome, diarrhoea,  
CC constipation, cirrhosis, jaundice, and hepatic vein thrombosis; nervous  
CC disorders including, but are not limited to, Alzheimer's disease,  
CC amnesia, bipolar disorder, cataplexy, cerebral neoplasms, Down's  
CC syndrome, and dystonias; smooth muscle cell disorders including, but  
CC not limited to, angina, anaphylactic shock, arrhythmia, cardiovascular  
CC shock, migraine, and pheochromocytoma; musculoskeletal disorders  
CC including muscular dystrophy, central core disease, nemaline myopathy,  
CC centronuclear myopathy, lipid myopathy, inclusion body myositis,  
CC thyrotoxic myopathy, and ethanol myopathy; immunological disorders  
CC including AIDS, Addison's disease, adult respiratory distress  
CC syndrome, allergy, ankylosing spondylitis, amyloidosis, anaemia,  
CC asthma, atherosclerosis, autoimmune hemolytic anaemia, autoimmune  
CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's  
CC disease, atopic dermatitis, dermatomyositis, diabetes mellitus,  
CC emphysema, episodic lymphopenia with lymphocytotoxicity, erythroblastosis  
CC fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,  
CC Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis,  
CC hyperesinophilia, irritable bowel syndrome, multiple sclerosis,  
CC myasthenia gravis, myocardial or pericardial inflammation,  
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,  
CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's  
CC syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic  
CC sclerosis, thrombocytopenic purpura, ulcerative colitis, urethritis,  
CC Werner syndrome, complications of cancer, haemodialysis, and  
CC extracorporeal circulation; viral, fungal, bacterial or protozoan  
CC infections; trauma; and neoplastic disorders including adenocarcinoma,  
CC leukemia, melanoma, myeloma, sarcoma, and various cancers. The REC  
CC polynucleotide is a source of probes and primers which bind may be used  
CC to detect REC in a sample from a patient (claimed). They may also be  
CC administered as part of a gene therapy regime.

XX  
SQ Sequence 1208 BP; 280 A; 323 C; 355 G; 250 T; 0 other;

Query Match 18.5%; Score 120.6; DB 21; Length 1208;

Best Local Similarity 59.0%; Pred. No. 5e-28;

Matches 207; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 228 TGTGACCTCTATACAGTGAAGTGAATGCTGCTGAGCAACAGATTGGTATCCCGT 287



XX	22-Oct-2001 (first entry)
DE	
XX	Human polynucleotide SEQ ID NO 4401.
XX	
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM	peripheral nervous system; neuropathy; central nervous system; CNS;
KM	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KM	leukaemia; ss.
OS	
XX	Homo sapiens.
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	
PA	(HYSE-) HYSEQ INC.
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AL, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
DR	WPI: 2001-442253/47.
XX	P-PsDB; AAM41256.
PT	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
PS	Claim 1; SEQ ID NO 4401; 10078pp; English.
CC	
CC	The invention relates to human nucleic acids (AA157798-AA161369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: immune system suppression,
CC	Activ/inhibit activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 1699 BP; 403 A; 459 C; 461 G; 376 T; 0 other;
Query Match	18.5%; Score 120.6; DB 22; Length 1699;
Best Local Similarity	59.0%; Pred.No.5,9e-28;
Matches 207; Conservative	0; Mismatches 144; Indels 0; Gaps 0;

QY	348	CACATTAGCCAAAGATTGCTGGTGGGTTATGTCATATATGATGATATAGTAAGAGC	407
Db	1275	CACCTGTGGGGAAGGCGACGATGGCATCTGTACACACTTGCTTACTCGGTCATCGGGC	1216
QY	408	ATGGATTATATACCAACCGTGTGTTGTGACCGCGAGTGAACACTTGATGTGGAGCAA	467
Db	1215	CTGGGACCGCCACCAACAGCCCTGCTCTTTCGGCCGGCCGATGAAACCGCATGTGGAGCA	1156
QY	468	TCCTTTCACAGAACGGCACCTGTTGCTTGTATATGAACCTTGGAAATCACCCTATTCCTCC	527
Db	1155	CCCGATACACAGCGGACAGTGAACACGCTCAAGGCTTGGCTATGTCGACATGCCCTG	1096
QY	528	CATCAAGAAAGAACTGGCCTGTGAGACACGATGATGGCGCAATGGCTGA	578
Db	1095	TGTGGCCAGAAAGCTGTGTCTGCGAGATGAAGTCTCGGGCCATGCGCTGA	1045
RESULT 11			
AA158626			
ID	AA158626	standard; cDNA; 1748 BP.	
XX	AA158626;		
XX	22-OCT-2001	(first entry)	
XX	Human polynucleotide SEQ ID NO 829.		
XX	Human; neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;		
XX	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia; ss.		
XX	Homo sapiens.		
OS	WO200153312-A1.		
PN	26-JUL-2001.		
XX	26-DEC-2000;	2000WO-US34263.	
XX	21-JAN-2000;	2000US-0488725.	
XX	25-APR-2000;	2000US-0552317.	
PR	09-JUL-2000;	2000US-0598042.	
PR	19-JUL-2000;	2000US-0620312.	
PR	03-AUG-2000;	2000US-0653450.	
PR	14-SEP-2000;	2000US-0662191.	
PR	19-OCT-2000;	2000US-0693036.	
PR	29-NOV-2000;	2000US-0727344.	
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
XX	Pang Z, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
PI	WPT: 2001-442253/47.		
DR	P-PsDB; AAM39470.		
XX	Novel nucleic acids and polypeptides, useful for treating disorders		
XX	such as central nervous system injuries -		
PS	Claim 1; SEQ ID NO 829; 10078bp; English.		
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AAM38642-AA42213) with neurotrophic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		







**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2002, 18:30:25 ; Search time 3143 Seconds

(without alignments)  
6046.494 Million cell updates/sec

Title: US-09-868-300-7

Perfect score: 653  
Sequence: 1 gaattcgacagcagctcctt.....caagaagatggaaccagttg 653

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- 1: gb\_da:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_da:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_hugo\_hum:\*
- 40: em\_hugo\_mus:\*
- 41: em\_hugo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653	100.0	653	6 AX027396	AX027396 Sequence
2	530.6	81.3	2522	8 AT080192	U80192 Arabidopsis
3	530.6	81.3	95295	8 AC020889	AC020889 Genomic s
4	316.2	48.4	1867	8 AF166262	AF166262 Arabidops
5	316.2	48.4	6588	8 AB026641	AB026641 Arabidops
6	167.6	25.7	288539	14 AF198100	AF198100 Fowlpox v
7	166.8	25.5	138906	2 AP003946	AP003946 Oryza sat
8	166.2	25.5	631	8 PAB271130	PAB271130 Picea abi
9	120.6	18.5	1167	9 AF182419	AF182419 Homo sapi
10	120.6	18.5	1240	9 BC014409	BC014409 Homo sapi
11	120.6	18.5	2211	9 AK027491	AK027491 Homo sapi
12	113.8	17.4	1074	3 AY102691	AY102691 Drosophila
13	100.8	15.4	12855	8 SPAC15E1	SPAC15E1 Saccharomyc
14	99.8	15.3	2206	8 SCYKR072C	SCYKR072C Caenorhabdi
15	99.8	15.3	4775	8 SCU01878	SCU01878 Streptococ
16	92.8	14.2	2187	8 CTHAL3GEN	CTHAL3GEN Trypanoso
17	92.8	14.2	155007	2 AC084047	AC084047 S.cerevisia
18	85.2	13.0	3022	8 SCYOR053W	SCYOR053W Streptococ
19	85.2	13.0	54719	8 SCXY55XB	SCXY55XB Streptococ
20	83.6	12.8	22239	3 CEF25H9	CEF25H9 Streptococ
21	81.2	12.4	10324	1 AE006562	AE006562 Streptococ
22	80.6	12.3	2145	8 SCYKL086W	SCYKL086W Streptococ
23	79.6	12.2	10324	1 AE010042	AE010042 Streptococ
24	79.6	12.2	50511	1 AE014154	AE014154 Streptococ
25	72.6	11.1	895	11 CNS06K5R	CNS06K5R Mus muscu
26	71.8	11.0	1678	10 BC004779	BC004779 Mus muscu
27	70.4	10.8	12975	1 AE006291	AE006291 Lactococc
28	69.2	10.6	10029	1 AE008484	AE008484 Streptococ
29	69.2	10.6	13002	1 AE007423	AE007423 Streptococ
30	69.2	10.6	77743	2 SPNEU1910	SPNEU1910 Streptococ
31	67.2	10.3	437	1 SMU48885	SMU48885 Streptococ
32	64.6	9.9	210185	2 AC125321	AC125321 Mus muscu
33	63.6	9.7	9909	6 BD003692	BD003692 Polynucle
34	62.8	9.6	165158	3 AC009912	AC009912 Drosophila
35	62.8	9.6	166146	2 AC009732	AC009732 Drosophila
36	62.8	9.6	167447	3 AC009459	AC009459 Drosophila
37	62.8	9.6	198244	3 AC007803	AC007803 Drosophila
38	62.8	9.6	313634	3 AE003454	AE003454 Drosophila
39	61.4	9.4	13249	1 AE002516	AE002516 Neisseria
40	61.4	9.4	349980	6 AX044033	AX044033 Sequence
41	58.2	8.9	326301	1 NMA622491	AL162757 Neisseria
42	57.2	8.8	10049	1 AE007681	AE007681 Clostridi
43	56	8.6	10066	1 AE010581	AE010581 Fusobacte
44	55.6	8.5	347400	1 AP003591	AP003591 Nostoc sp
45	55.2	8.5	110000	2 LMFLCHR32_26	Continuation (27 o

ALIGNMENTS

RESULT 1	LOCUS	SEQUENCE	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE
AX027396	AX027396	653 bp	DNA	Linear	PAT 16-SEP-2000		Arabidopsis thaliana			
AX027396	AX027396	7 from Patent WO0036124.					Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi			
AX027396.1	GI:10188367						thale cress.	1 (bases 1 to 653)	de Veylder,L., Inze,D., Boudolf,V.K. and Torres,A.J.	Novel cell cycle genes and uses thereof

JOURNAL	Patent: WO 0036124-A 7 22-JUN-2000; VEYDER LIEVEN DE (BE) ; INZE DIRK (BE) ; CROPPDISIGN N V (BE) ; BOUDOLF VERONIQUE KATELLIJNE CE (BE) ; TORRES ACOSTA JUAN ANTONIO (BE)
FEATURES	Location/Qualifiers 1..653 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" <1..>651 /note="unnamed protein product" /codon_start=1 /protein_id="CAC09138.1" /db_xref="GI:10188368" /translation="EFGTSSFLGCKNIEKKMNEVDVTVRKRPILLAAAGSVASIKFS NICHSEMAEVAIVASRSLNPFVDRPSLEQNVTLTYTDEDEWSNMKIGDPLVHIELR RMADVMIIAPLNTLAKIAGLIDNLICIVRAMDYSKPLFVAPAMNTLMNNPTE RRLVLDLGITLIPPIKRLACGDYNGAMERPSLIVSRFLFWSOARKORDGTS"
CDS	
BASE COUNT	180 a 135 c 151 g 187 t
ORIGIN	
Query Match	100.0%; Score 653; DB 6; Length 653; Best Local Similarity 100.0%; Pred. No. 5.9e-187; Matches 653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GAATGGGACAGACCTGCTCCGCTGATACAAAGATAGAGAGATGATATGAA 60 
Db	1 GAATGGGACAGACCTGCTCCGCTGATACAAAGATAGAGAGATGATATGAA 60 
OY	61 GTGATACAGTACAGAGAAAGCCTGATCTTACTAGCTGCAAGTGTGGCTTCA 120 
Db	61 GTGATACAGTACAGAGAAAGCCTGATCTTACTAGCTGCAAGTGTGGCTTCA 120 
OY	121 ATTAAGTTCAGTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180 
Db	121 ATTAAGTTCAGTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180 
OY	181 TCAGATATCATCTCATATTTGCTGATTAACCTTCTCTACTCATGATGACTCTCAT 240 
Db	181 TCAGATATCATCTCATATTTGCTGATTAACCTTCTCTACTCATGATGACTCTCAT 240 
OY	241 ACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300 
Db	241 ACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300 
OY	301 CTCGAGCGCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 360 
Db	301 CTCGAGCGCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 360 
OY	361 ATTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420 
Db	361 ATTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420 
OY	421 AAACCGTTGTTTGTGCAACCGGCGATGAACACTTTGATGTGAACATCTTTTCACAGAA 480 
Db	421 AAACCGTTGTTTGTGCAACCGGCGATGAACACTTTGATGTGAACATCTTTTCACAGAA 480 
OY	481 CGGACACCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540 
Db	481 CGGACACCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540 
OY	541 CTGCGCTGTGAGAGACTACGCTAATGCGCAATGGCTGACCTTCTCTGATTTATTCACACT 600 
Db	541 CTGCGCTGTGAGAGACTACGCTAATGCGCAATGGCTGACCTTCTCTGATTTATTCACACT 600 
OY	601 GTTAGACTGTTCTGGAGTCAAGCTGTAACAAGAGATGACACCACTTG 653 
Db	601 GTTAGACTGTTCTGGAGTCAAGCTGTAACAAGAGATGACACCACTTG 653 
RESULT 2	
LOCUS	ATU080192 2522 bp DNA linear PLN 04-MAY-2001
DEFINITION	Arabidopsis thaliana HAL3B protein (HAL3B) gene, complete cds.

QY	87	TATCTTACAGTCGACGAGTGAAGTGTGGCTTCGAATTAAGTTCACATTAATCTCTGCCATTG	146
Db	1353	TATCTTACAGTCGACGAGTGAAGTGTGGCTTCGAATTAAGTTCACATTAATCTCTGCCATTG	1412
QY	147	TTTCTCAGAAATGGGTGAAGTCAAAAGCCGTGCTTCAAAATCATCTCTCATTTTGCTTGA	206
Db	1413	TTTCTCAGAAATGGGTGAAGTCAAAAGCCGTGCTTCAAAATCATCTCTCATTTTGCTTGA	1472
QY	207	TAAACCTTCTCTACCTCAGAAATGTGACTCTCTATACAGTGAAGTGAATGGTCTAGCTG	266
Db	1473	TAAACCTTCTCTACCTCAGAAATGTGACTCTCTATACAGTGAAGTGAATGGTCTAGCTG	1532
QY	267	GAACAAGATTTGGTGATGCCGTTCTTCATATGAGCTCAGACGCTGGGCTGATGTTATGAT	326
Db	1533	GAACAAGATTTGGTGATGCCGTTCTTCATATGAGCTCAGACGCTGGGCTGATGTTATGAT	1592
QY	327	CATTGCTCTTTTGTCTGTCTTAACACATTAGCCA-----	358
Db	1553	CATTGCTCTTTTGTCTGTCTTAACACATTAGCCAAGGCTACTAATATATAGAAACGTAGC	1652
QY	359	-----AGATTGCT	366
Db	1653	TTTAAATCTGGATTTTACTACTATTGGCTTCATATTGTTTTTTTGTGTGGAGATTGCT	1712
QY	367	GCTGGCTTATGTGATTAATCTATTGACATGTATAGTAAGCATGGGATTAATGACAAACCG	426
Db	1713	GCTGGCTTATGTGATTAATCTATTGACATGTATAGTAAGCATGGGATTAATGACAAACCG	1772
QY	427	TTGTTTGTGGACCGGGCATGAACCTTGATGTGAGAACATCTTTACAGAACGGGAC	486
Db	1773	TTGTTTGTGGACCGGGCATGAACCTTGATGTGAGAACATCTTTACAGAACGGGAC	1832
QY	487	CTTGTCTTCTTGATGAACCTTGATGAATCAACCCCTAATCTCCCATCAAGAACATGTGGCC	546
Db	1833	CTTGTCTTCTTGATGAACCTTGATGAATCAACCCCTAATCTCCCATCAAGAACATGTGGCC	1892
QY	547	TCTGAGACATACGGTAATATGGCGCAATGGCTGAGCCTTCTTGATTTATTCACACTGTAGA	606
Db	1893	TCTGAGACATACGGTAATATGGCGCAATGGCTGAGCCTTCTTGATTTATTCACACTGTAGA	1952
QY	607	CTGTCTCTGGAGTCAACAGCTCTGTAAACAAAGAGATGGAACACAGTTG	653
Db	1953	CTGTCTCTGGAGTCAACAGCTCTGTAAACAAAGAGATGGAACACAGTTG	1999
RESULT 3			
AC020889/c		95295 bp	DNA linear
LOCUS			PLN 28-JUN-2000
DEFINITION			Genomic sequence for Arabidopsis thaliana BAC T1N15 from chromosome
ACCESSION			I. complete sequence.
VERSION			AC020889
KEYWORDS			AC020889.3 GI:6938141
SOURCE			HMG.
ORGANISM			Arabidopsis thaliana.
			Arabidopsis thaliana.
			Eukaryotes: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
			Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
			Kosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE			1 (bases 1 to 95295)
AUTHORS			Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C.,
			Shim, P., Altafi, H., Bei, Q., Chin, C., Chion, J., Choi, E., Conn, L.,
			Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B.,
			Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukherjee, N.,
			Muylven, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A.,
			Thaler, A., Tortum, M., Vaysberg, M., Yu, G., Federspiel, N.A.,
			Genomic sequence for Arabidopsis thaliana BAC T1N15 from chromosome
TITLE			I.
JOURNAL			Unpublished
REFERENCE			2 (bases 1 to 95295)
AUTHORS			Ecker, J. R.
TITLE			Direct Submission

JOURNAL	Submitted (12-JAN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE	3 (bases 1 to 95295)
AUTHORS	Ecker,J.R.
TITLE	Direct Submission
JOURNAL	Submitted (08-FEB-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE	4 (bases 1 to 95295)
AUTHORS	Chen,K.R., Shim,P., Brooks,S., Buchler,E., Chao,O., Johnson-Hopson,C., Khan,S., Kim,C., Alfati,H., Bel,B., Chin,C., Chion,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howling,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukhasky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE	Direct Submission
JOURNAL	Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT	On Feb 8, 2000 this sequence version replaced gi:6693011.
FEATURES	Location/Qualifiers
SOURCE	1..95295
CDS	/organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" /clone="TIN15" complement(join(<1..384,461..506)) /note="Hypothetical protein" /codon_start=1 /evidence=not_experimental /product="TIN15.1" /protein_id="AAF79719.1" /db_xref="gi:8778711"
CDS	/translation="WDASQPLPQLPSLSDESRNQVRSALDNDTEGFSRAEM F/RASPLSLSDCDDYATILDRPNPNPOVIPHDDTDITSEESTLSPQME VSTCMOKNIEDLVSVLTANLEKTAHLIQRROTLOEA" complement(join(1177..1388,1453..1524,1626..1720, 1799..1872,1960..2104,2188..2251,2337..2392,2463.. 2759..3018,3107..3269,3358..3480,3831..3902,4001.. 4577..4693,4765..4903,4994..5100,5194..5303,5387.. 5603..5776,5832..6061,6149..6792,7208..7313)) /note="similar to zwille protein emb CAA11429.1; similar to ESTs gb AA7207232.1, dbj AV442417.1, gb A1993691.1, dbj AV41187.1, and gb AA651494.1" /codon_start=1 /evidence=not_experimental /product="TIN15.2" /protein_id="AAF79718.1" /db_xref="gi:8778710"
CDS	/translation="MAAVLYFLHLICVRSAYELDFSTRIGSSAPFCYIMRKRR TPABEGGGSSRACGYSGGGRSGRGQQGGGGRQGGRGTTPPOQGGRGRT GPDPQOOQYGPOEYQGRGPRPHGGRGGGGRGGSPPORQOSVELHQAT MPTGVASSOPLSEVSPOVEPVVAIQOEOLSEVGAFSAOIOPIPSSSKAKRP SBRYAGOSGKRCIVKANHFPAELPKDLAHVDITPEVTSGYNRAVMKOLDNYND SHLGSRYPADGRKSIFYTAGLPFPNSKEFRINLLDEEVGAGGSSSYGLREBEFY VIKVARADIHLGMLEFGKSDAPQEDLOYLDIRLPIGRXWILLIETRIYIPG NSTPYBDIGKKOSLEDGLESMRGTYOSTIRPOMGLSIGVACYSMTASTEAPYI QPVCOLLNRDISRLSDADRARIKRALRGVAEVTHINGNNRRYRISGLAAVATREL TPVEDVNTOKSVSEFHEHYGFRIIOHTOLPCLOGNSNPARYIPMEVCKIVEGORS KLNRNEOTALLKVTCORPIDRENKILOTLOYLNDAKONYKOEEGIKISTISLAEVR ILPFWLIYHESGREGTCLPOVGONMMNKMGITNGVYNNMICINFPSNOVDNLARFE COELAQMYCSGMARNPREPVLPYASREOVEXUKLTRYHDATJKTSOGKETDLLIYI LIDNNKSLTGDKRICETEFELIVSOCCLRKHVFMSKQMANVAALKINVKGNNTYL VDALSRLPLVSDRPLIFEGADVTHPHGEDSSPSIAAVASQMPETIYAGLCVQO AIRCOLIDOLFKEMLDPOKGVATGMIKYVMMLFNINFTIGELLTAFRSTGHKLRII FPRDVSGEOFFVOVLLEYLDAIKRACALEGQTOPVTVYVCAKRHHTRLEAQNHRN BSVDNSGNILGETSTJVDKSICTPDTEDELFCGHAIGOGSRPAHYHVLMDENPFA DGLQSLTNLCCTYAKCTRYSYIHPRAYIALAARAFIYPEPTSOGSSMASSMAR GGGMASTRGRNGNAARPLPALKENKRWMTYC" complement(join(11659..11863,11978..12075,12232..12384, 12425..12482,12570..12662,12789..12880,12982..13092,
CDS	

```
13176..13436,13522..13695))
/note="similar to 1-aminocyclopropane-1-carboxylate
deaminase (ACC) gb|AA05069.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.3"
/protein_id="AAF79717.1"
/db_xref="GI:8778709"
/translation="MGRSLTISRYKLEIARSMATSVPSNADELITTKPYSPSNAS
HLRPLSHFSLAHLPTPIHWNLPDLPNGTSLWKRDQFTGMEISLGNKVRLEPLMA
EVDVADHADVITIGTIOGNHCRATATASNVLNLSLILRTSKLADPPRGVGNLIL
EVLGANNHLISKEEYSISSEALFNALKEKLEKRGKPYVIPVGGSNLGTWYIEA
AARIEQNLNRPDPLDKFDIVVACGGGTAGISGMYGALKAKLTDSVKEPFYVO
VIAFSYCDPDEYFVPGVCGLLDGHAGVNSRIVNHNNAKGKGYAMNTSEELFVKY
AASGVYIDPVSIGRAITCLINETTKDPKCMESKRLFTHTGSLGLYKIVQDMASLM
GMSRMDSVSESVKDGVGKMF"
/complement(join(17766..17891,18000..18095,18178..18290,
18375..18475,18551..18634,18739..18779,18849..18974,
19123..19196,19300..19595,19752..19807,19958..20041,
20151..20324,20522..20611,20677..20754,20864..20911,
21000..21112,21193..21256,21333..21406,21542..21581))
/note="similar to dihydroxyacetone kinase
sp|P4510|DHAK_CITPR; similar to ESTs gb|T42731.1,
gb|AA042761.1, and gb|N96033.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.4"
/protein_id="AAF9716.1"
/db_xref="GI:8778708"
/translation="MYPARKFTINPNVDVTEFEGLVETYPGLQYLIDGLPELKVIAR
ADVSANYNKVAIVISGGSGHEPQAGYVEGMLTAAICGVDFASPTVSIILARIAY
TGKGCILVNTYNTGDRNLGFLAEOAKGDEIETVTVIVGDDCALPPRGVARGRG
LACTVYVHYKAGAAAAGSLAEVAAEAKHASMGTMGVALSVTCALPGQVSDRILG
EKRELGLVHGEPGAAVVDIQPEVYVSHVLOQILSPETNYVTVITGNSVYIMNIG
GTPLEMLTAAKGAVALDIEGLAVDRYVTSFMTSLDAGCCFCCVIESLITLCA
LILPFGSFTSIMKADOSILERLDATNPANSPVPTDLSRSPAKIPVLPERS
TKREEESOOGRILEAAIEAAATVINKLNMPOKGVDDGCGSGMCGAVALIED
MKRYPLMDAETYSIEGSSIRRYMGTSGLIHPTVSEYETYNLCAAYAEELANFS
ETTSKHSFALKSISASVSKYGATGCTGATGCTMADALIPASVLEEKISVGEDPIAAEVL
SAFAATAGASTIHKRQAGRSSYSVAEIFASPDPGMAAAAMYSAAAAVAVEQROR
F"
join(21967..22138,22278..22378)
/note="unknown protein; similar to ESTs gb|AF162845.1,
gb|AI992553.1, and gb|AA042117.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.5"
/protein_id="AAF79693.1"
/db_xref="GI:8778685"
/translation="MALOWLISVVAVEVTVLVLTPPYMLLKKRVOLVLSLIDP
AASIVAFAGROLDIYKAEHLRLSCSEVCTATPRDRYEKSV7IP"
/complement(join(23638..24481,24564..24859,25106..25237))
/note="unknown protein; similar to EST gb|AI996045.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.6"
/protein_id="AAF79715.1"
/db_xref="GI:8778707"
/translation="MDAVANLFOSPSLKQTSRLFTSPNCGFMRISGPOFRIR
ONSLSKFSRPFOSGTTCKSRSFVAKASAGDSTSIYALQKTSVGGOLSLDILVS
HPHLVPAVEQOLEQOLIDRADEOSDASVLETQDILVRLAEVKEKREYEMLEBI
LYLVQKFMNDANTVLPSTISSADSPGADVPTDIDGELERASPEVEYMIONHLS
IILKNRPDQDITAVNOTSKIGOVYASVWYGYLKRIDOFOLKETMRILIPGSDSG
ETISLEQGRDVERNFTEAEETTYAASVNDVGSFVGINASGFSDDMSOKRKYV
MSEDEYTLQRAITRSRESVGLIEHTEALGREGVITTPQGITDSSKDEHITKISFG
LKRVLAVYTFGSLMDVESHVDSRYHFLV"
join(26361..26450,26546..26588,26676..27014,27366..27706,
27823..28041)
/note="unknown protein; similar to ESTs emb|Z33851.1,
gb|T33009.1, and gb|AA394999.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.7"
13176..13436,13522..13695))
/note="similar to 1-aminocyclopropane-1-carboxylate
deaminase (ACC) gb|AA05069.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.3"
/protein_id="AAF79717.1"
/db_xref="GI:8778709"
/translation="MGRSLTISRYKLEIARSMATSVPSNADELITTKPYSPSNAS
HLRPLSHFSLAHLPTPIHWNLPDLPNGTSLWKRDQFTGMEISLGNKVRLEPLMA
EVDVADHADVITIGTIOGNHCRATATASNVLNLSLILRTSKLADPPRGVGNLIL
EVLGANNHLISKEEYSISSEALFNALKEKLEKRGKPYVIPVGGSNLGTWYIEA
AARIEQNLNRPDPLDKFDIVVACGGGTAGISGMYGALKAKLTDSVKEPFYVO
VIAFSYCDPDEYFVPGVCGLLDGHAGVNSRIVNHNNAKGKGYAMNTSEELFVKY
AASGVYIDPVSIGRAITCLINETTKDPKCMESKRLFTHTGSLGLYKIVQDMASLM
GMSRMDSVSESVKDGVGKMF"
/complement(join(17766..17891,18000..18095,18178..18290,
18375..18475,18551..18634,18739..18779,18849..18974,
19123..19196,19300..19595,19752..19807,19958..20041,
20151..20324,20522..20611,20677..20754,20864..20911,
21000..21112,21193..21256,21333..21406,21542..21581))
/note="similar to dihydroxyacetone kinase
sp|P4510|DHAK_CITPR; similar to ESTs gb|T42731.1,
gb|AA042761.1, and gb|N96033.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.4"
/protein_id="AAF9716.1"
/db_xref="GI:8778708"
/translation="MYPARKFTINPNVDVTEFEGLVETYPGLQYLIDGLPELKVIAR
ADVSANYNKVAIVISGGSGHEPQAGYVEGMLTAAICGVDFASPTVSIILARIAY
TGKGCILVNTYNTGDRNLGFLAEOAKGDEIETVTVIVGDDCALPPRGVARGRG
LACTVYVHYKAGAAAAGSLAEVAAEAKHASMGTMGVALSVTCALPGQVSDRILG
EKRELGLVHGEPGAAVVDIQPEVYVSHVLOQILSPETNYVTVITGNSVYIMNIG
GTPLEMLTAAKGAVALDIEGLAVDRYVTSFMTSLDAGCCFCCVIESLITLCA
LILPFGSFTSIMKADOSILERLDATNPANSPVPTDLSRSPAKIPVLPERS
TKREEESOOGRILEAAIEAAATVINKLNMPOKGVDDGCGSGMCGAVALIED
MKRYPLMDAETYSIEGSSIRRYMGTSGLIHPTVSEYETYNLCAAYAEELANFS
ETTSKHSFALKSISASVSKYGATGCTGATGCTMADALIPASVLEEKISVGEDPIAAEVL
SAFAATAGASTIHKRQAGRSSYSVAEIFASPDPGMAAAAMYSAAAAVAVEQROR
F"
join(21967..22138,22278..22378)
/note="unknown protein; similar to ESTs gb|AF162845.1,
gb|AI992553.1, and gb|AA042117.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.5"
/protein_id="AAF79693.1"
/db_xref="GI:8778685"
/translation="MALOWLISVVAVEVTVLVLTPPYMLLKKRVOLVLSLIDP
AASIVAFAGROLDIYKAEHLRLSCSEVCTATPRDRYEKSV7IP"
/complement(join(23638..24481,24564..24859,25106..25237))
/note="unknown protein; similar to EST gb|AI996045.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.6"
/protein_id="AAF79715.1"
/db_xref="GI:8778707"
/translation="MDAVANLFOSPSLKQTSRLFTSPNCGFMRISGPOFRIR
ONSLSKFSRPFOSGTTCKSRSFVAKASAGDSTSIYALQKTSVGGOLSLDILVS
HPHLVPAVEQOLEQOLIDRADEOSDASVLETQDILVRLAEVKEKREYEMLEBI
LYLVQKFMNDANTVLPSTISSADSPGADVPTDIDGELERASPEVEYMIONHLS
IILKNRPDQDITAVNOTSKIGOVYASVWYGYLKRIDOFOLKETMRILIPGSDSG
ETISLEQGRDVERNFTEAEETTYAASVNDVGSFVGINASGFSDDMSOKRKYV
MSEDEYTLQRAITRSRESVGLIEHTEALGREGVITTPQGITDSSKDEHITKISFG
LKRVLAVYTFGSLMDVESHVDSRYHFLV"
join(26361..26450,26546..26588,26676..27014,27366..27706,
27823..28041)
/note="unknown protein; similar to ESTs emb|Z33851.1,
gb|T33009.1, and gb|AA394999.1"
/codon_start=1
/evidence=not_experimental
/product="T1N15.7"
Query Match 81.3%; Score 530.6; DB 8; Length 95295;
Best Local Similarity 88.1%; Pred. No. 1,4e-149;
Matches 623; Conservative 0; Mismatches 4; Indels 80; Gaps 1;
CDS
27 CTGTAAACAAGTATAGAGAAAGATGATATGGAAGTATGAGTAACTAACAGAAAGCCCTCG 86
|||
|||
|||
Db 84994 CTGATTTCAGTATAGAGAAAGATGATATGGAAGTATGAGTAACTAACAGAAAGCCCTCG 84935
87 TATCTTACTAGCTCAAGTGGAGTGGCTGGCTTCATTAAGTTCAAGTAATCTCGCATTCG 146
TATCTTACTAGCTCAAGTGGAGTGGCTGGCTTCATTAAGTTCAAGTAATCTCGCATTCG
Db 84934 TATCTTACTAGCTCAAGTGGAGTGGCTGGCTTCATTAAGTTCAAGTAATCTCGCATTCG 84875
147 TTTCTCAGATGGCGCTGAGTCAAAAGCCGCTGCTTCAAAATCATCTCATATTGCTGTA 206
TTTCTCAGATGGCGCTGAGTCAAAAGCCGCTGCTTCAAAATCATCTCATATTGCTGTA
Db 84874 TTTCTCAGATGGCGCTGAGTCAAAAGCCGCTGCTTCAAAATCATCTCATATTGCTGTA 84815
207 TTAACCTTCTCTACCTCAGATGAGTGTGACTCTCTATACAGATGAAGATGAATGCTCTAGCTG 266
TTAACCTTCTCTACCTCAGATGAGTGTGACTCTCTATACAGATGAAGATGAATGCTCTAGCTG
Db 84814 TTAACCTTCTCTACCTCAGATGAGTGTGACTCTCTATACAGATGAAGATGAATGCTCTAGCTG 84755
267 GAACAAGATTGGTATCCGCTTCTTCATATGAGCTCAGACGCTGGCGTGATGTTATGAT 326
GAACAAGATTGGTATCCGCTTCTTCATATGAGCTCAGACGCTGGCGTGATGTTATGAT
Db 84754 GAACAAGATTGGTATCCGCTTCTTCATATGAGCTCAGACGCTGGCGTGATGTTATGAT 84695
327 CATTTGCTCTTTGCTGTCTAACAATTAGCCA ----- 358
|||||
|||||
|||||
Db 84694 CATTTGCTCTTTGCTGTCTAACAATTAGCCAAGGTACTAATTAATCATTAGCAAAACGTAGC 84635
359 -----AGATTGCT 366
|||||
|||||
|||||
Db 84634 TTTAGATCTGCATTTTACCTTTTGGCTTCAATATTTGTTTTTTTGTGTTGCAGATTGCT 84575
367 GGTGGGTATGTGAATATCTATTGACATGTATAGTAAGAGCATGGATTTATAGCAAAACCG 426
|||||
|||||
|||||
Db 84574 GGTGGGTATGTGAATATCTATTGACATGTATAGTAAGAGCATGGATTTATAGCAAAACCG 84515
427 TTGTTTGTGACACGGCGCATGAACTTTGATGTGGAACATCTTTTACGAACGGAC 486
TTGTTTGTGACACGGCGCATGAACTTTGATGTGGAACATCTTTTACGAACGGAC
Db 84514 TTGTTTGTGACACGGCGCATGAACTTTGATGTGGAACATCTTTTACGAACGGAC 84455
487 CTGTCCTGCTGATGATGAACTGGAATCAACCTTAATTCCTCCATCAAGAAACTGGCC 546
CTGTCCTGCTGATGATGAACTGGAATCAACCTTAATTCCTCCATCAAGAAACTGGCC
Db 84454 CTGTCCTGCTGATGATGAACTGGAATCAACCTTAATTCCTCCATCAAGAAACTGGCC 84395
```

QY	547	TGTGGAGACTACGGTAATAGCGCGCAATGGCTGAGCTTCTGTGATTTATTCACACTGTAGA	606
Db	84394	TGTGGAGACTACGGTAATAGCGCGCAATGGCTGAGCTTCTGTGATTTATTCACACTGTAGA	84335
QY	607	CTGTTCCTGGAGTCAACAGCTGTAAACAAGATGGAACCACTTG	653
Db	84334	CTGTTCCTGGAGTCAACAGCTGTAAACAAGATGGAACCACTTG	84288
RESULT 4			
LOCUS	AF166262	1867 bp	DNA linear
DEFINITION	Arabidopsis thaliana HAL3A protein (HAL3A) gene, complete cds.		
ACCESSION	AF166262		
VERSION	AF166262.1	GI:5802224	
KEYWORDS			
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 1867)		
TITLE	Espinoza-Ruiz, A., Belles, J. M., Serrano, R. and Cullanez-Macia, F. A. Arabidopsis thaliana HAL3A: a flavoprotein related to salt and osmotic tolerance and plant growth		
JOURNAL	Plant J. 20 (5), 529-539 (1999)		
MEDLINE	20117504		
PUBMED	10652125		
REFERENCE	2 (bases 1 to 1867)		
AUTHORS	Cullanez-Macia, F. A., Espinoza-Ruiz, A. and Serrano, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-JUL-1999) Instituto de Biologia Molecular y Celular de Plantas (UPV-CSIC), Camino de Vera s/n, Valencia 46022, Spain		
FEATURES	Location/Qualifiers		
source	1..1867		
gene	/organism="Arabidopsis thaliana"		
mrna	/db_xref="taxon:3702"		
cds	<954..>1700		
	/gene="HAL3A"		
	join(<954..1289,1407..>1700)		
	/gene="HAL3A"		
	/product="HAL3A protein"		
	join(954..1289,1407..1700)		
	/gene="HAL3A"		
	/note="Similar to Saccharomyces cerevisiae HAL3"		
	/codon_start=1		
	/product="HAL3A protein"		
	/protein_id="AA051616.1"		
	/db_xref="GI:5802225"		
	/translation="MENGKRLDROMENVTPRRKRVLLAAGSVAIIFKGLCHCTFTWAEKRVVTKSSLHFLDKLSLPQEVTLITDDESSMKKIDDPVHLERWADLVLVLAAPLSANLGLKLIAGGLCNLNLICILIRADYDKPLFVAAPAMNTLMNPPETERNLLSLDELGTLTLPPIKRRACGSDYGNGAMAEPSLIVSTVRLFESEAHOOTGTS"		
BASE COUNT	535 a 381 c 380 g 571 t		
ORIGIN			
Query Match	48.48;	Score 316.2;	DB 8; Length 1867;
Best Local Similarity	70.08;	Pre. No. 1.1e-84;	
Matches 502;	Conservative 0;	Mismatches 98;	Indels 117; Gaps
QY	53	ATATGGAAGTGTGATACAGTACAAGAGAGCGCTCGATCTTACTAGCTCAAGTGAAGTG	112
Db	982	ACATAGGAAGTGAATATCCACACCGAGAGAGCCTCGTACTACTCGCTCAAGTGAAGTG	1041
QY	113	TGCGTTCAATTAGCTAGTAATCTCTGCCATTTGTTTCTCAGATGGGCTGAAGTCAAG	172
Db	1042	TGCGTCTATCAATTAATTCGGCAATCTCTGCCATTTGTTTCTCAGATGGGCTGAAGTCAAG	1101
QY	173	CCGTCGGTCAAAATCAATCTCTCAATTTTCGTGAATAAACCTCTCTACTCGAATGGA	232
Db	1102	CCGTCGGTCAAAATCAATCTCTCTACTCTCTGATAAACCTCTCTCTCCACAGAAGTGA	1161

QY	233	CTCTCTATACAGATGAATGAATGCTTACTGTGAACAAGATTGGTATCCGGTCTTC	292
Db	1162	CTCTGTATACGATGATGAATGATATGCTTACTGTGAACAAGATCGTATCTGTCTTC	1221
QY	293	ATATCGAGCTCAGACGCTGGCGCTGATGTTATGATCATTTGCTCTCTTGTCTTAACACAT	352
Db	1222	ACATCTGACCTTAGACGTTGGGCTGATGTTATTAAGTCATTGCTCTTGTCTGTAAACCT	1281
QY	353	TAGCCA 	358
Db	1282	TAGCCAGAGTATTTAACCAATTCGACGAAACCCTAACCAACTACTAGATTCAATTGT	1341
QY	359	-----	358
Db	1342	GTGACTTGTGAACATTCATGAGAGATTTCATTCATCAATATATTTTGTTCACCTT	1401
QY	359	---AGATTGCGTGGGGGTATGTGATATTCATTAAGATATAGTATAGTAGGGGATT	415
Db	1402	TGTGATTTGCTGCTGGGGCTTTGTGATATATCTTCTGACATTATTAACGAGCTGGGGACT	1461
QY	416	ATAGCAAAACCGTTGTTGGTGCACCGGGGATGAACACTTTGATGTGAACAATCCTTTCA	475
Db	1462	ATACCAACCACTGTTTGTGCTCCAGCTATGAATACTTTGATGTGAACAATCCTTTCA	1521
QY	476	CAGAACGGCACCCTGTCTGCTGATGAACCTTGAATACCCCTAATTCCTCCATCAGA	535
Db	1522	CTGAAGGCACTTTTGTCTCTTGAATGAACGGGAGATCACACTTATTCCTCTTACCAAGA	1581
QY	536	AGAAATGGCCTGTGGAGACTACGGTAATGCGCAATGGCTGAGAGCCTTCTGTATTTATT	595
Db	1582	AGAGACCTTGCTGTGGAGACTACGGTAATGAGAGCTATGGTGAGCCCTCTCTATCATATT	1641
QY	596	CCACTGTTAAGCTGTTCTGGGAGTACACAAGCTCGTAACCAAGAAGATGGAAACCACTT	652
Db	1642	CCACTGTGAAGCTGTTCTGGGAGTCTCAGGCTCATACACAACCGGTGAACCTACTTT	1698

LOCUS	AB026641	6588 bp	DNA	linear	PLN 27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, Pl clone: MBG14.				
ACCESSION	AB026641 BA000014				
VERSION	AB026641.1 GI:4757397				
KEYWORDS					
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_1lb:Mitsui Pl clone:MBG14.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryotes: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae: eucosids II; Brassicales; Brassicaceae; Arabidopsids.				
REFERENCE	1 (sites)				
AUTHORS	Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.				
TITLE	Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the 4,251,695 bp regions covered by 90 Pl, TAC and BAC clones				
JOURNAL	DNA Res. 7 (3), 217-221 (2000)				
MEDLINE	2036309				
REFERENCE	2 (bases 1 to 6588)				
AUTHORS	Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1552-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yasukamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)				
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see <a href="http://www.kazusa.or.jp/kaos/cgi-bin/gad_graph.cgi?c=MBG14">http://www.kazusa.or.jp/kaos/cgi-bin/gad_graph.cgi?c=MBG14</a> Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <a href="http://compbio.ornl.gov/Grail-1.3/">http://compbio.ornl.gov/Grail-1.3/</a> ),				



GB:AF021350"  
/codon\_start=1  
/product="ORF FPV001"  
/protein\_id="AAf44603.1"  
/db\_xref="GI:7271757"  
/translation="MKMNPIDIEYKEKMKFLRDNSEYILPITCLTSVNTSLCPA  
ALFVAVNDCKRDSFLEDTTATITSSSIATYRNLYIHCPRDMSHNGICLSTGK  
VGRROGIQKCEKLGSDMIGKSEEMKALKINMGNDISRCFVNDRAAASFDPVNECA  
YGRSSSYVEPKVLTPSCSVRRYLVCNTKNTSYPTTQSSFFNYE"  
3367..4035  
/gene="FPV002"  
3367..4035  
/note="identical to FPV259 Fowlpox GB:A06621"  
/gene="FPV002"  
/codon\_start=1  
/product="ORF FPV002 hypothetical protein"  
/protein\_id="AAf44597.1"  
/db\_xref="GI:7271751"  
/translation="MIVEKIAAMLYPLCLRCICNSVPRATCKCYHCLLYPEVCC  
ECMSERTDSEHSCCTCCVPLILIRFRRVILPTLKATDCIRLPCVLTRPKRT  
ICPLASWCRCFCPCCEVFLRCLFPCMLRRMRGRLTGVRERGARDSRDPARGT  
WVNDWCEDLCVMIWSPCCYVRCIRTCMTFTTKIFYWTAPAGSPHMPPEPPLSRK  
VFSS"  
complement(4500..4871)  
/gene="FPV003"  
complement(4500..4871)  
/gene="FPV003"  
/note="identical to FPV258 Fowlpox PIR:C31685 (vaccinia  
A40R), similar to Homo sapiens SW:Q07108"  
/codon\_start=1  
/product="ORF FPV003 C-type lectin gene family protein"  
/protein\_id="AAf44606.1"  
/db\_xref="GI:7271760"  
/translation="MKSLLIAMLMDCANSINCGRPYTSYNKCIWNRDLKMHK  
KTYSEASTCLTFEPMTLARSLIDNEKDMKFISKGMGSLMRIDDKKPEVCKAY  
TDGKTGFSPCNATYGFVCIID"  
5125..5430  
/gene="FPV004"  
5125..5430  
/note="identical to FPV257 Fowlpox GB:A06621"  
/gene="FPV004"  
/codon\_start=1  
/product="ORF FPV004 hypothetical protein"  
/protein\_id="AAf44598.1"  
/db\_xref="GI:7271752"  
/translation="MNNDITFTLFYCKNKRYGEGRRGKGTILFHPINRVICT  
SARQCKYTRIGIKIYAVARHVSITRCGRSHARVDRKPSFGQKDFHICAGSA"  
complement(5221..5589)  
/gene="FPV005"  
complement(5221..5589)  
/gene="FPV005"  
/note="identical to FPV256 Fowlpox PIR:E31685"  
/codon\_start=1  
/product="ORF FPV005 ERG gene family protein"  
/protein\_id="AAf44604.1"  
/db\_xref="GI:7271758"  
/translation="MALNLRVRPEDESGIKVLAASDPLNYENDHTGDCFYARAR  
NMEFCRARYLCAAGDTVKIYFLEGELIYSRVGSPADSGYVTRGNCVEFETDS  
SCFTLMCTSSYNTVYVYVME"  
5931..7187  
/gene="FPV006"  
5931..7187  
/gene="FPV006"  
/note="identical to FPV255 Fowlpox SW:P14361 (vaccinia  
C10L), similar to SW:P03296"  
/codon\_start=1  
/product="ORF FPV006 C4L/C10L-like gene family protein"  
/protein\_id="AAf44589.1"  
/db\_xref="GI:7271753"  
/translation="MEFSCGTGNNQNLAVHRTETRTSFKEKLLVNLGISDLNDIKNI  
CEGSKTFEPKRETLISIKDRKSQIYFENSLNDLKLKHALIYDELSTVDNVAYVE  
NWTLLMYEKGDFARHRDSVSKNIICVHLLIYIPEPTEGGEIVYIYIDNNTSVKL  
KTHLFDKTIHESITIVESGRKCVALFDVILLEKLSASTNVIGSIEITLCKINLYDRE

NDLQICDVIEMRTEDKEYSIGMISDRSGRCIKSHHNGSIYVRKEEYSPDALCI  
YNNAEVDEIWTGDKKHIITWSTIDKTKTSTFIPIDPVLYEKIKAISSKEHEYDRLRF  
CNSRTEYICCSVSKYFPDLPTKTDLHEVINSIDYDRKSVGTPDWYTLPLEVQFTILG  
NMSYEELFNIVRGMIALEEDNEVGC"  
7599..7817  
/gene="FPV007"  
7599..7817  
/note="identical to FPV254 Fowlpox GB:D00295"  
/codon\_start=1  
/product="ORF FPV007 hypothetical protein"  
/protein\_id="AAf44600.1"  
/db\_xref="GI:7271754"  
/translation="MLYSVFTRIYKIRKNNKTEMIATYDRNAVYKKGISGILLRD  
NGTRNMLNITIDFSSRNQMVLKVSVL"  
7681..8184  
/gene="FPV008"  
7681..8184  
/gene="FPV008"  
/note="identical to FPV253 SW:P14370 (vaccinia A40R),  
similar to Gallus gallus GB:AB015628"  
/codon\_start=1  
/product="ORF FPV008 C-type lectin gene family protein"  
/protein\_id="AAf44607.1"  
/db\_xref="GI:7271761"  
/translation="MPLILRKQVSESCVAITVIGLICLILFTLVVVTCKWYAPY  
FSKVCPEDEWIGNSKCYFTINETNMDSKLCDDVMSLSIRFNIETLNFVSYYGSG  
SYWIDIQNRKIRPQINFSLYEGGVNDICLFTPSNIEMSCIFHERTICVKEDRYTH  
WYEPYMR"  
8048..8248  
/gene="FPV009"  
8048..8248  
/gene="FPV009"  
/note="identical to FPV252 Fowlpox GB:D00295"  
/codon\_start=1  
/product="ORF FPV009 hypothetical protein"  
/protein\_id="AAf44601.1"  
/db\_xref="GI:7271755"  
/translation="MKNALMIFVYLYLTRVLTSLKCLVYFTKERYVLKIDIPICIPNTC  
VRFSTFLYSILYLVNRSKY"  
complement(9123..10190)  
/gene="FPV010"  
complement(9123..10190)  
/note="Fowlpox SW:P14369, similar to Homo sapiens  
GB:AB006423"  
/codon\_start=1  
/product="ORF FPV010 Serpin gene family protein"  
/protein\_id="AAf44354.1"  
/db\_xref="GI:7271508"  
/translation="MGSILVRLKLELYVPGKDCISPGVYTIILMIMIGCKKETRDKI  
KDLIGFNGVVPPIPKSEYVEYVDDELINISIMIEGYPKRRPFIKIDPDIENA  
KYVSGPDDTSETINKVVELSTRGLIKDESISLADRIALINIVLYKSKWKRPDE  
LTSKHPKRYNGTDVMDITMDIDNAVYTHDEDRSVOVLEEDREFWFIIPDS  
VTSIGDVDSLNGKNKINIKIISKDMTLKTIYVLPFELEDDVDLKDALIHGCCNDL  
FKSGELVIGDFTKRLRIGNIRORSVIRKVEYGEAASVETSCDTDGKIKPIVKANPV  
FMFLVAVQTKPIELFGLIFOG"  
complement(10276..11112)  
/gene="FPV011"  
complement(10276..11112)  
/gene="FPV011"  
/note="similar to Bos taurus PIR:S32367"  
/codon\_start=1  
/product="ORF FPV011 alpha-SNAP"  
/protein\_id="AAf44355.1"  
/db\_xref="GI:7271509"  
/translation="MOKIKLVSDFEKNLFGPIETIEBAAGVSHAANLFAVQLMBC  
AGKAFKSGDMLQKNKNSIAAASSFVDANAERKIDSYBAINCLSAILEVYCLKRF  
YTVARCKMINIAIYENDILELDRKAIFHEYNAASGYGEGYKNSLSDCMLIARLSIOK  
EDDRDRAKIEIOGVYNNMMLSKYESRHOLLVLAIMCYLCSVDSRAKRSIDKXDLFP  
AFKDFKCKRIEKILAACERKNIEFTSALIEEDYDHGNTIDEALMSMLLTIRKATFDE  
VE"  
complement(12038..13033)





Accession	Sequence	Position
Dp_113058	TCTTGAGACATCAACCTGCTAGGATATCTTTGGTCCCTCCCATTAACCAAAAGCGTGGC	113117
Qy_546	CTGTCGAGACTACGCGTATGCGCGCAATGGCTGAGCGCTTCTGATTTATTCCACTGTTAG	605
Db_113118	CTGTGCTCATATATGGTATATGGTCGCAATGGCTGAGCGCTTCTGTATCGATATCCACCGTCAG	113177
Qy_606	ACTGTTCGAGATCACAGGCTCGTAAACAAA	637
Dp_113178	GCTTGCTTGCAAGACAGCCACTTAATACAA	113209

RESULT 8	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
PAB271130	PAB271130	631 bp mRNA linear Picea abies partial mRNA for putative halotolerance protein HAL3 homolog, (PpA0038 gene).	AJ271130	AJ271130.1	GI:12583570	halotolerance protein HAL3 homolog; PpA0038 gene. Norway spruce.	Picea abies

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1	Schubert, R., Mueller-Starck, G. and Riegel, R.	Development of EST-PCR markers and monitoring their intrapopulational genetic variation in <i>Picea abies</i> (L.) Karst	Theor. Appl. Genet. 103 (8), 1223-1231 (2001)	2 (bases 1 to 611)

**JOURNAL** Submitted (24-JAN-2000) Dr. Schubert R., Faculty of Forest Sciences, Section of Forest Genetics, Technical University of Munich, Am Hochanger 13, D-85354 Freising, GERMANY

**FEATURES** \_location/Qualifiers

FEATURES	location/qualifiers
source	1. .631

gene	1. .363
CDS	/gene="pPA0038" <1. .363

BASE COUNT ORIGIN	174 a	135 c	126 g	195 t	1 others
----------------------	-------	-------	-------	-------	----------

Query Match	25.5%;	Score 166.2;	DB 8;	Length 631;
Best Local Similarity	68.9%;	Pred. No. 3.4e-39;		
Matches 228;	Conservative 0;	Mismatches 103;	Indels 0;	Gaps

280 GATCCGTTCTTCATATCGAGCTCAGACGCTGGGCTGATGTTAATGATCATTTGCTCCTTG 33

Db 10 GGTCTGTACTGCATATAGACTCCGGCATGGGCTGATGCTATGGTATTGCTCCACTA 69

340 1C10C1ACACAI1AC0CAB0AI1G1G1G1G0BI1A1G1G1A1AC1I1AT1G1ACAI1ATA 35

[illegible]

Db 130 ATACGTCATGGGACCTTTAAACAAGCCCTCTTTTGTAGCTCCGCAATGAATACCTTCATG 18

QY	460	TGGAACATCCCTTCACAGAAAGGACCTGTCTGCTGATGAATCGGAATACACCTA	519
Db	190	TGGAACACCCATTACTCAAGGCACTTTGGACTCTATTCGGAGATGGAGATACACTT	249
QY	520	ATTCCTCCCATCAAGAGAAACTGGCCCTGTGGAGACTACGGTAAATGGCGCAATGGCTGAG	579
Db	250	ATCCCCCTATTAACAAGAAGCGTTAGCTTGTGGTCAATTATGGAAATGGTGCAMATGTCAGAA	309
QY	580	CCCTCTCGATTATTAATCCACATGGTGAAGCTGT	610
Db	310	CTTCTCCATATGATACAACTCTTAGGTTTT	340

RESULT	9			PRI	20-SEP-2000
LOCUS	AF182419	1167 bp	mRNA	linear	
DEFINITION	Homo sapiens MDS018 (MDS018)	mrna,	complete cds.		
ACCESSION	AF182419				
VERSION	AF182419.1	GI:10197637			
KEYWORDS	.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 1167)	Huang, C., Qian, B., Tu, Y., Gu, W., Wang, Y., Han, Z. and Chen, Z.	Novel genes expressed in hematopoietic stem/progenitor cells from myelodysplastic Syndromes patient

REFERENCE  
2 (bases 1 to 1167)  
AUTHORS Huang, C., Qian, B., Tu, Y., Gu, W., Wang, Y., Han, Z. and Chen, Z.  
TITLE Direct Submission  
JOURNAL Submitted (02-SEP-1999) Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, People's Republic of China

FEATURES	Location/Qualifiers
source	1. .1167

```

CDS      /gene="MDS018"
         252..635
         /gene="MDS018"

```

BASE COUNT	274 a	318 c	338 g	237 t
AMAEVGTIVDKYKEVLFQHSFGQSS"				

Query Match	18.5%	Score 120.6;	DB 9;	Length 1167;
Best Local Similarity	59.0%;			
Matches 207; Conservative	0;	Mismatches 144;	Indels 0;	Gaps

228 TGTGACTCTCTATACAGATGAAGATGATGCTTACGCTGGAACAGATTGGTATCCCGT 287

Db 218 TCTCACCCCTCTACAGCGACGCTGATGATGGAGATGTGGAAGACCGCTCTGACCCAGT 277

288 TCTTCATATCGAGCTCAGACGCTGGGCTGATGTTATGATCATTCCTCCTTTGTCGCTAA 347

D5 2/8 1CTGCACATTGACCTGCGGAGGTGGGGCAGACCCTCTGCTGTGGCTCCTCTTGAAGCCAA 33/

[illegible]

408 ATGGGATTATAGCAAAACCGTTGTTTGTTCACCGGCGATGAACCTTTGATGTGGAAACAA 467

Db 398 CTGGACCGCAGCAAGCCCTGCTCTTCCGCCGCATGAACCCCATGTGGAGCA 457

Qy 468 TCCTTACAGAGAGCGACCTGTCTGTGTGATGACCTGAATACCCATATCCCTCC 527

Db 458 CCCGATCACAGCGCAGCGAGTGTAGACCTCAAGGCTTTGGCTATGTGTGAGATCCCCG 517

Qy 528 CATCAAGAAGAACTGGCTGTGTGAGACTACGCTAATGCGCCGAATGGCTGA 578

Db 518 TGTGGCCAAAGAGCTGTGTGTGGAGATGAAGTCTCGGGGGCCATGGCTGA 568

RESULT 10

BC014409 1240 bp mRNA linear PRI 19-SEP-2001

LOCUS Homo sapiens, hypothetical protein MDS018, clone MGC:19897

DEFINITION IMAGE:4647763, mRNA, complete cds.

ACCESSION BC014409.1 GI:15680132

VERSION BC014409.1

KEYWORDS MGC.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1240)

AUTHORS Strausberg, R.

JOURNAL Direct Submission

Submitted (17-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: [cgaabs-remail.nih.gov](mailto:cgaabs-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

[info@cgsc.bc.ca](mailto:info@cgsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosche, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letitia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Seedei, Jacqueline Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stolt, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Zaira.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 29 Row: n Column: 19

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 11141896.

Location/Qualifiers

1. 1240

/organism="Homo sapiens"

/db\_xref="locusid:60490"

/db\_xref="taxon:9606"

/clone="MGC:19897 IMAGE:4647763"

/tissue\_type="Eye, retinoblastoma"

/clone\_lib="NIH\_MGC\_16"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

71. 685

/codon\_start=1

/product="hypothetical protein MDS018"

/protein\_id="AAH14409.1"

/db\_xref="GI:15680133"

/translation="MEPRASCPAAAPLMERKFFHYLVGVTSVAALKPLIVSKLLIDP

GLEAVYVTERAKHVFSPDIPVLYISPADDEWMSKSRDPVLIIDLRRADLLIAP

LDANTLKVASGICDNLITCYMRAMDSKPLIFCPANMTAMENPITRAOVDOLKARF

YVEIPVAKTKLVCGDESGIAGAAVEGTVDRKVEVLFQHSFGQSS"

CDS

BASE COUNT 313 a 321 c 356 g 250 t

ORIGIN

Query Match 18.5%; Score 120.6; DB 9; Length 1240;

Best Local Similarity 59.0%; Pred. No. 2.5e-25;

Matches 207; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

Qy 228 TGTGACTCTATPACAGATGAAGATGAATGTCTAGCTGGAACAAGATTGATCCCGT 287

Db 268 TGTCAACCCCTACAGCAGCAGCTGATGAATGGAGATGGGAAGAGCCGCTGACCCAGT 327

Qy 288 TCTTATATCAGAGCTCAGACGCTGGCTGATGTATGATCATCTGCTCTTCTGCTAA 347

Db 328 TGTGCACATTGTGACCTGCGAGGTGGCAGACCTCTCTGTGGTGGCTCTTGTGAAGCCA 387

Qy 348 CACATTAGCAAGATTGTGCTGGTGGTGTATGTATGATATGACATGTATGATGAAGC 407

Db 388 CACTCTGGGGAGAGTGGCCAGTGCATCTGTGACAACTTGCTTACTTGCTGATCGGGGC 447

Qy 408 ATGGATTATACCAACCGTTGTTGTTGACCGCGCATGAACACTTTGATGTGAGACAA 467

Db 448 CTGGACCGCAGCAGACCCCTGCTCTTGTGCGCCGCATGAACACCGCATGTGGAGACA 507

Qy 468 TCCTTACAGAGAGCGACCTTGTCTGCTTGTATGATGATGATGATGATGATGATGATG 527

Db 508 CCCGATCACAGCGCAGCAGTGTAGACCACTCAAGGCTTGTGTATGTGTGAGATCCCTG 567

Qy 528 CATCAAGAAGAACTGGCTGTGTGAGACTACGCTAATGCGCCGAATGGCTGA 578

Db 568 TGTGGCCAAAGAGCTGTGTGTGGAGATGAAGTCTCGGGGGCCATGGCTGA 618

RESULT 11

AK027491

LOCUS Homo sapiens CDNA FLJ14585 fis, clone NT2RM4001611, weakly similar to S1S2 PROTEIN.

DEFINITION AK027491.1 GI:14042205

ACCESSION AK027491

VERSION AK027491.1

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens teratocarcinoma cell line:NT2 CDNA to mRNA, clone\_lib:NT2RM4 clone:NT2RM4001611.

ORGANISM Homo sapiens

REFERENCE 1

AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosofiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahata,K., Masuko,Y., Ninomiya,K. and Iwayanagi,T.

TITLE NEBO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2211)

AUTHORS Isogai,T. and Otsuki,T.

JOURNAL Direct Submission

TITLE Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 297-0812, Japan (E-mail:genom@celshri.co.jp, Tel:01-438-52-3975, Fax:01-438-52-3986)

COMMENT NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

Location/Qualifiers

1. 2211

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="NT2RM4001611"

/cell\_line="NT2"

FEATURES

SOURCE



KEYWORDS acetyltransferase complex subunit; actin-interacting protein; ardi family; fat1; glutaredoxin; phosphoethanolamine cytidyltransferase; ribosomal protein l44; thioltransferase; thymidylate synthase-like DNA metabolism protein; vacuolar sorting protein; yeast bud6/Alp3p homolog.

SOURCE Schizosaccharomyces pombe.

ORGANISM Schizosaccharomyces pombe Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.

REFERENCE 1 (bases 1 to 12855) Murphy, L., Harris, D., Barrell, B.G., Rajandream, M.A. and Lyne, M.H. Direct Submission Submitted (09-AUG-1999) Schizosaccharomyces pombe chromosome I sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ E-mail: barrell@sanger.ac.uk

COMMENT Notes: Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/projects/S\\_pombe](http://www.sanger.ac.uk/projects/S_pombe))  
Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in Pombase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.  
CDS are numbered using the following system eg SPAC5H10.01c, SP (S. pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid c15E1 is overlapped at the 5' end by cosmid c15A10.  
Location/Qualifiers  
1. 12855  
/organism="Schizosaccharomyces pombe"  
/strain="g72h-"  
/db\_xref="taxon:4896"  
/chromosome="I"  
/map="ICEN"  
/clone="cosmid c15E1"  
1. 1093  
/gene="SPAC15E1.01"  
/note="SPAC15A10.15; fat1"  
1. 1093  
/partial  
/gene="fat1"  
/note="SPAC15E1.01, len:>363"  
/codon\_start=2  
/label=fat1  
/product="yeast actin-interacting protein Bud6/Alp3p homolog"  
/protein\_id="CAB52420.1"  
/db\_xref="GI:5725406"  
/translation="NVEKFEDEVNDQIELMRSDVLRKVRGDLQDOTSOLNEESAVLEKRIQTLERLEEVPTIMKQWERELNAIVQDEFLDSHTVLISDLKRDLSALSTVLSNYSALIELAKRSIKSKPLTKATESEIIGRDQIQLQEVNLINPKPSEARLQALIESSELOKRIQLRQVDEFSKEVKTVEENEKLNKGAEADRIITIDEKVYKTIIMDPYSSKRNKNGSFIESSPTVIDEHYVDPNSAKATVAELIDYGSQVTENPMLEPSLAIKPKLKKPNTIYETIVYSTAHETDEQTPSKYSINRVSSSDTYFENTDLKIDNNVMSKYTHVHRHDTISTDDYDAEDADVEETSLT"  
1. 35  
/gene="SPAC15E1.01"  
/note="nominal overlap with cosmid SPAC15A10, EM:297208 S. pombe chromosome 1"  
complement(1440..1886)  
/gene="SPAC15E1.02c"  
complement(1440..1886)

/gene="SPAC15E1.02c"  
/note="SPAC15E1.02c, len:148"  
/codon\_start=1  
/label=SPAC15E1.02c  
/product="hypothetical protein"  
/protein\_id="CAB52421.1"  
/db\_xref="GI:5725407"  
/translation="MPSYHMPYVRPSAILGIAFDHNAFLVYGPLMGDIWRKAMSTAPHVEYDAGSKTKKVAATYVNSFLSTAVOSYSIALLOLTGTVLKGFVGLYVFGASGLPDVDYMFTESSRGTPYILVKTISLSVYKGVSLVIGVARI"  
join(2639..2938,3129..3149)  
/gene="SPAC15E1.03"  
join(2639..2938,3129..3149)  
/gene="SPAC15E1.03"  
/note="SPAC15E1.03, len:106, SIMILARITY:phaffia rhodozyma, RL44\_P4ARH, 60s ribosomal protein l44, (105 aa), fasta scores: opt: 608, E():0, (81.98 identity in 105 aa)"  
/codon\_start=1  
/label=SPAC15E1.03  
/product="60s ribosomal protein l44"  
/protein\_id="CAB52422.1"  
/db\_xref="GI:5725408"  
/translation="MWNIPKTKTYCGEKNCRKTVYRKYQKKGPSKLAQGRRD RKQSGFGQTRPVFHKRAKRVKVVRLQVSCYKKNQVLKRCRHFELGEEKTKGA A10P"  
2693..2923  
/gene="SPAC15E1.03"  
/note="Match to PF00935 Ribosomal\_L44, Ribosomal protein L44, Score 174.98"  
2939..3128  
/gene="SPAC15E1.03"  
/note="confirmed intron"  
2939..2944  
/gene="SPAC15E1.03"  
/note="gtatgt, splice donor sequence"  
3113..3128  
/gene="SPAC15E1.03"  
/note="ctacaacatcttag, splice branch and acceptor"  
3963..5840  
/gene="SPAC15E1.04"  
3963..5840  
/gene="SPAC15E1.04"  
/note="SPAC15E1.04, len:625, N-term SIMILARITY:Candida tropicalis, SIS2\_CANTR, SIS2 PROTEIN (HALOTOLERANCE PROTEIN HAL3), (531 aa), fasta scores: opt: 613, E(): 1.3e-29, C-term SIMILARITY:Saccharomyces cerevisiae, TYSY\_YRST, thymidylate synthase, (304 aa), fasta scores: opt: 1403, E():0, (65.7% identity in 303 aa)"  
/codon\_start=1  
/label=SPAC15E1.04  
/product="thymidylate synthase-like DNA metabolism protein"  
/protein\_id="CAB52423.1"  
/db\_xref="GI:5725409"  
/translation="MSQPLHARFAFRAVKNPMLKEKROLTDSKYHILVAATGSVAIKLTLIVSLTYKGVQVVLTPDARNEVEKEDLTALGVVYNNADDMKWDGIECPITHIELRRMHLILAPLSANTMAKMGICNLITSLIRAWAPLKPIILAPMTLEKMTNPTTQHLISASIRYKNSEFIMPILEVACGDIQGMGMEMRIIVGRVADKIQLEKSVLEPNAYKNIDGDDDSSEQTAEEYEDDDDDVDVNDNEOSNIEISANADITPKASLLPSTTESISKDHETSQAPLGSESVPTQASENVTTKPEPVPTTSSEYRTEEQYLNLIRYILENGSGRPDRGTGTRSVFAPQPLFSLRNNTLPILTKRVFLRGVLEELIMFIHDDYKGNVHLSEKGIHIMDNGSREELDSGLDRVYGDIGPYGQMHHFGQAYVDCDTYTKNGVQLAOVSTILKNYDRITLISANPLAIPMALPCOHICQAFVSPCKGCKGROLSMNYORADMGICVPMNIASISLTTHMTAHMCGYEAAPFVHMGDCHITNDLHLAQTOLEHVPKAFKPLFKRKDAKDIGSIDPSVDDFAVEGNTPIGPIKMKSV"  
4944..5837  
/gene="SPAC15E1.04"  
/note="Match to PF00303 thymidylat\_synth, Thymidylate synthase, Score 685.78"  
complement(join(6052..7082,7138..7204))  
/gene="SPAC15E1.05c"  
complement(join(6052..7082,7138..7204))

	/gene="SPAC15E1.05c"	
	/note="SPAC15E1.05c, len:365, SIMILARITY:Homo sapiens,	
	/99447, phosphoethanolamine cytidylyltransferase., (389	
	aa), fasta scores: opt: 713, E():0, (37.0% identity in 357	
	aa)"	
	/codon_start=1	
	/label="SPAC15E1.05c	
	/product="phosphoethanolamine cytidylyltransferase (EC	
	2.7.7.14)"	
	/protein_id="CAB52424.1"	
	/translation="MASSSNIKRRMLMDGCMDFRHYGSNAILIQAOLGETLYGHIS	
	DEELTLKKGPVMTLEERCSLANTCKWDEVPSAPYFPLEWMRRGGCYVHGDI	
	STANDGDYCAFAKAADQYLEVKRTEDSVTELLDRLLSPLEIVSTPVSLSQID	
	LIRRFATDSDLGFPTFDVFYINTEKPETLLSGTLLLNENKNITIIDGMDELTEKH	
	ISALETCTRMPPGPIIMAGIFPADCEPEKPMNLLENLIINLOCKYSITLVGPAPS	
	LPASSKYIKLCFPDIOISKVVYPVPSTPVSLPALDISLNPNNNSFYFDKLGSLLKQ	
	RVMRRHREERQRKRKKNADETDTTKTTA"	
	complement{join(6727..7082,7138..7171)}	
	/gene="SPAC15E1.05c"	
	/note="Match to PF01467 Cytidylyltransf,	
	Cytidylyltransferase score 117.35"	
	complement(7083..7096)	
	/gene="SPAC15E1.05c"	
	/note="ctaacactcatg, splice branch and acceptor"	
	complement(7132..7137)	
	/gene="SPAC15E1.05c"	
	/note="tagga, splice donor sequence"	
	join(7593..7655,7734..8234)	
	/gene="SPAC15E1.06"	
	join(7593..7655,7734..8234)	
	/gene="SPAC15E1.06"	
	/note="SPAC15E1.06, len:187, SIMILARITY:Arabidopsis	
	thaliana, CAB41864, hypothetical 21.0 kb protein., (190	
	aa), fasta scores: opt: 609, E():0, (45.7% identity in 186	
	aa)"	
	/codon_start=1	
	/label="SPAC15E1.06	
	/product="similar to yeast vacuolar sorting protein	
	Vps29/PEP1"	
	/protein_id="CAB52425.1"	
	/db_xref="GI:5725411"	
	/translation="MLVLVTIGDEFHIIPRARLKSEFRQLIIPKIISOILCLNTYSTS	
	VVEYLKHCSDCLKLVKGAFDISSKAPIAGKITGSFKIGYNGLVLPDOSPALSLI	
	AAREDDADILLFEGGNHKFAAYELDCCFPGVNGSATGARNVASVEDEKIVPSVLMVDQ	
	GAVLLVYRIIPGEVREVMQYRKPF"	
	7656..7661	
	/gene="SPAC15E1.06"	
	/note="tagga, splice donor sequence"	
	7712..7733	
misc_feature		
Query Match	15.4%:	Score 100.8; DB 8; Length 12855;
Best Local Similarity	55.6%:	Pred. No.3.e-19;
Matches 239;	Conservative 0;	Mismatches 182; Indels 9; Gaps 2;
QY	158 GGGGTGAAGTAAGAAGCGTCGCCCTTCACAATAATCTCTCAATTGCCTTGATAAACCTTCTC	217
Db	4135 GCGTGCAGCTTCAGTGCCTTTTAAACGAGATCGCGCTCGTAACCTTGTGTAGAAAAGAGACT	4134
QY	218 TA--CCTCAGAATGTGACTCTCATACAGATGAAGTAAGTAAGTGTCTAGCTGGAACAAGA	274
Db	4195 TAAAGCGTTTGGGTGTCATGTTTATACATCGGAGATGATTGGAAAAATTGGACGAC	4254
QY	275 TTGGTGATCCCGTCTTCATATCGAGCTCAAGCGCTGGGCTGATGTTATGATCTGCTC	334
Db	4255 TCGAATGTCCCATACACATATTTAGACTGCTGATGGGCAATCTTTGTTAATTGGCC	4314
QY	335 CTTTGCTGTCTAACACATTAACCAAGATTGCTGGTGGTTATGTGATATCTATTGACAT	394
Db	4315 GTTTAAGTGCAAAATATCTATGCGCAAAATGGCAATGCTCTGTGACACATTTACTACTCT	4374
QY	395 GTTATGAGAGCATGGATTATAGCAAAACCGTTGTTGTTGACCGCGCATGAACACTT	454

DB	Accession	Gene	Protein	Length	Score	DB	Length	Score	DB	Length	Score
Db	4375	CTCTTATACGTGCTCGGCCCCCTTTAAACCAATTCTCTTGCGCTCCCGCAATGAATACCT		4434							
Oy	4455	TGATGTGGAACATCTCTTACACAGAACGACACTGTCTTGCTTGATGAACCTTGGATCA		514							
Db	4435	TGATGTGGAACATCTCTTACACAGAACATTTTGAGCGCTATTAGCAATATACAGAA		4494							
Oy	515	-----CCCTAAATTCCTCCATCAAGAAAGAAACCTGCGCTGTGAGACATCAGGTAAATGGCG		568							
Db	4495	ATAGATTAATTCATGATCGCAATCGAAGAAAGTGCTTGCTGATGATATATGATGATGGCG		4554							
Oy	569	CAATGGCTGA		578							
Db	4555	GTAATGGCGGA		4564							
RESULT 14											
LOCUS	SCYKR072C		2206 bp	DNA		linear		PLN 11-AUG-1997			
DEFINITION	S. cerevisiae chromosome XI reading frame ORF YKR072C.										
ACCESSION	Z28297 Y13137										
VERSION	Z28297.1 GI:486544										
KEYWORDS											
SOURCE											
ORGANISM	Saccharomyces cerevisiae.										
REFERENCE	Saccharomyces cerevisiae										
AUTHORS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;										
AUTHORS	Saccharomycetales; Saccharomycetaceae; Saccharomyces.										
REFERENCE	1 (bases 1 to 2206)										
TITLE	Pohl,T.M. and Pohl,F.M.										
JOURNAL	Unpublished										
AUTHORS	2 (bases 1 to 2206)										
JOURNAL	MIPS.										
DEFINITION	Submitted (09-MAY-1994) Data collected by MIPS on behalf of the										
DEFINITION	European yeast chromosome XI sequencing project. MIPS at the										
DEFINITION	Max-Planck-Institut fuer Biochemie, Am Klopfersplitz 18a D-82152										
DEFINITION	Martinsried, FRG; E-mail: Mewes@helmholtz-mips.biochem.mpg.de										
DEFINITION	Location/Qualifiers										
FEATURES											
source	1..2206										
	/organism="Saccharomyces cerevisiae"										
	/db_xref="taxon:4932"										
	/chromosome="XI"										
	/complement(455..2143)										
	/gene="SIS2"										
	complement(455..2143)										
	/gene="SIS2"										
	/note="ORF YKR072c"										
	/codon_start=1										
	/protein_id="CAA82151.1"										
	/db_xref="GI:486545"										
	/translation="MTAAVASTGSGKODAHNOSIECPFSRSGOKELILDHEDAGKOSI										
	INSEVSGROSTISPLSNATTTTTSKINATGTSAAVSNTPEROLKRVAVTSDLKO										
	QOKODSLTLLKNDSEKTSPPNSNPAPVSNIPGHAIVPHNTSRTQLSGSPVNE										
	MKDIDTPKLDKNSALKIVDTMKPDKIMATSTPISRNKNVTAAPTSILRKEDQODAN										
	NVSQINIVSTPEETPVKOSVIPSIIIPKRNKSLDPLRPDDGKLHLVGAGSISLV										
	EKIPMKRLKEETIYGRDRIQVLIQSAOFPEQRTKIKISSEKLNKMSQYESFP										





Db 63 TWKKITPDPVLIHLELRMDVFIAPLNTANTLAKIANGICDNLITSVIRAWDTNKPILFC 122

QY 147 PAMNTLMNNPTEHRLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSSLISTVR 202

DB 123 PAMNTLMNNPTEHRLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSSLISTVR 178

## RESULT 2

QY 09AVS8 PRELIMINARY: PRT: 120 AA.

AC 09AVS8:

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Putative halotolerance protein HAL3 homolog (Fragment).

GN PPA0038.

OS Picea abies (Norway spruce) (Picea excelsa).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

OX NCBI\_TaxID=3329;

RN NCBI\_TaxID=3329;

RP SEQUENCE FROM N.A.

RC STRAIN=KARST;

RA Schubert R., Mueller-Starck G., Riegel R.;

RT "Development of EST-PCR markers and monitoring their intrapopulation

RL genetic variation in Picea abies (L.) Karstl.";

DR EMBL: AJ271130; CAC27336.1; -.

DR HSSP: Q9SWE5; 1E20

DR InterPro: IPR003382; Flavoprotein.

DR Pfam: PF02441; Flavoprotein; 1.

FT NON\_TER 1

SQ SEQUENCE 120 AA; 13064 MW; 4164223419A49FDA CRC64;

Query Match 41.5%; Score 475; DB 10; Length 120;

Best Local Similarity 79.4%; Pred. No. 5.1e-40;

Matches 85; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

QY 96 VLIHLELRMDVFIAPLNTANTLAKIANGICDNLITSVIRAWDTNKPILFC 155

DB 6 VLIHLELRMDVFIAPLNTANTLAKIANGICDNLITSVIRAWDTNKPILFC 65

QY 156 NPTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSSLISTVR 202

DB 66 NPTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSSLISTVR 112

## RESULT 3

QY 096SX0 PRELIMINARY: PRT: 204 AA.

AC 096SX0:

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE CDNA FLJ1585 fis, clone NT2RM4001611, weakly similar to SIS2

DE protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;

RN NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.

RA Isogai T., Nagai K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,

RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,

RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,

RA Ninomiya K., Iwayanagi T.;

RT "NDO human cDNA sequencing project";

RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AK027491; BAB5151.1; -.

DR InterPro: IPR003382; Flavoprotein.

DR Pfam: PF02441; Flavoprotein; 1.

SQ SEQUENCE 204 AA; 22429 MW; 98808FA0207C6B66 CRC64;

Query Match 33.9%; Score 445; DB 4; Length 204;

Best Local Similarity 48.4%; Pred. No. 1e-36;

Matches 90; Conservative 26; Mismatches 64; Indels 6; Gaps 3;

QY 26 RKPRILLASGSVASIKFSNLCHEFE--WAEYKAVASKSLNFVDKPSLPQNTVLTDE 83

DB 16 RKPRILLASGSVASIAALKLPLVSKLIDIPGLEVSVTTERAKHFSPDIP--VTLYSDA 73

QY 84 DEWSSWNKIGDPPVLIHLELRMDVFIAPLNTANTLAKIANGICDNLITSVIRAWDTNKP 143

DB 74 DEWEMKSSSDPVLHIDLRMDVFIAPLNTANTLAKIANGICDNLITCVIRAWDRSKPL 133

QY 144 EVAPAMNTLMNNPTEHRLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSSLISTVR- 202

DB 134 LFCPAMNTLMNNPTEHRLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSSLISTVR 193

QY 203 -LEWES 207

DB 194 VLFQHS 199

## RESULT 4

QY 096CD2 PRELIMINARY: PRT: 204 AA.

AC 096CD2:

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Hypothetical 22.4 kDa protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;

RN NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.

RC TISSUE=EYE;

RA Strusberg R.;

RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: BC014409; AAH14409.1; -.

DR InterPro: IPR003382; Flavoprotein.

DR Pfam: PF02441; Flavoprotein; 1.

KW Hypothetical protein.

SQ SEQUENCE 204 AA; 22413 MW; 9899683486C32A54 CRC64;

Query Match 38.8%; Score 444; DB 4; Length 204;

Best Local Similarity 48.4%; Pred. No. 1.3e-36;

Matches 90; Conservative 26; Mismatches 64; Indels 6; Gaps 3;

QY 26 RKPRILLASGSVASIKFSNLCHEFE--WAEYKAVASKSLNFVDKPSLPQNTVLTDE 83

DB 16 RKPRILLASGSVASIAALKLPLVSKLIDIPGLEVSVTTERAKHFSPDIP--VTLYSDA 73

QY 84 DEWSSWNKIGDPPVLIHLELRMDVFIAPLNTANTLAKIANGICDNLITSVIRAWDTNKP 143

DB 74 DEWEMKSSSDPVLHIDLRMDVFIAPLNTANTLAKIANGICDNLITCVIRAWDRSKPL 133

QY 144 EVAPAMNTLMNNPTEHRLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSSLISTVR- 202

DB 134 LFCPAMNTLMNNPTEHRLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSSLISTVR 193

QY 203 -LEWES 207

DB 194 VLFQHS 199

RESULT 5

QY 09UT17 PRELIMINARY: PRT: 625 AA.

AC 09UT17:

DT 01-MAY-2000 (TREMBLrel. 13, Created)



DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
DE Thymidylate synthase-like DNA metabolism protein.  
GN SPAC15E1.04.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxId=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972H-;  
RA Murphy L., Harris D., Barrell B.G., Rajadream M.A., Lyne M.H.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: 5,10-METHYLENETETRAHYDROFOLATE + DUMP  
CC -1- PATHWAY: DEOXYRIBONUCLEOTIDE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE SYNTHASE FAMILY.  
CC EMBL: AL009770; CAB52423.1; -.  
DR HSSP: P13100; IC17.  
DR InterPro: IPR003382; Flavoprotein.  
DR InterPro: IPR000398; Thymidylat\_synth.  
DR Pfam: PF02441; Flavoprotein; 1.  
DR Pfam: PF00303; thymidylat\_synth; 1.  
DR PRINTS: PR00108; THYDMSNTASE.  
DR ProDom: PD001180; thymidylat\_synth; 1.  
DR PROSITE, PS00091: THYMIDYLATE\_SYNTHASE; 1.  
KW Methyltransferase; Nucleic acid biosynthesis; Transferase.  
KW SEQUENCE 625 AA; 69993 MW; B3BC6034BF26F61E CRC64;

Query Match	36.8%;	Score 421.5;	DB 3;	Length 625;
Best Local Similarity	48.8%;	Pred. No. 9.8e-34;		
Matches 84;	Conservative 31;	Mismatches 52;	Indels 5;	Gaps 3;

[illegible]

RESULT 6  
P91988  
ID P91988 PRELIMINARY; PRT; 237 AA.

DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE F25H9.6 protein.  
GN  
NC Caenorhabditis elegans.  
OC  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCB1\_TaxID=6239;  
NM [1]  
NM  
RP SEQUENCE FROM N.A.  
RA Smye R.;  
RL submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology".  
RL Science 282:2012-2018(1998).  
DR EMBL; Z81069; CAB02993.1; -.  
DR HSSP; Q9SWE5; IE20.

DR InterPro: IPR003382; Flavoprotein.  
DR Pfam: PF02441; Flavoprotein\_1.  
SEQUENCE 237 AA: 26900 MW: 4A33F0A2620CB303 CRC64:

Query Match	36.8%;	Score 421;	DB 5;	Length 237;
Best Local Similarity	48.9%;	Pred. No. 3.2e-34;		
Matches 90;	Conservative 23;	Mismatches 59;	Indels 12;	Gaps 3;

Qy	27	KPRLILASGVASIKRPSNLCHESEME-----VKAVASKSINFPDKPSLPONT	78
Db	42	KNHLLILITISIAVKAPEL---ISELEYEKIGDRRIILKVYTTENAKKLCHIOKLEDEI	98
Qy	79	LYTDEDEWSSNNKIGDPVLHIELRRMADVMIAPISANTLAKITAGCLDNLITCIVAM	138
Db	99	YVEDDEDEWSSMRREKGDVLHIELRRMADSLIAPLDANTMAKINGCLDNLVYSIIAM	158
Qy	139	YSKPLEVAPAMNTLMNNPTEERH-LVLDLELGITLPIPKKILACGDYNGAMAEPSLI	197
Db	159	LSKPCYFAPAMNTLMNENPLTMOHRTVLKSOLKREKEICPIOKELICGDVGTGAMASIGTI	218
Qy	198	YSTV	201
Db	219	VSIV	222

RESULT 7  
Q9HEH6  
ID Q9HEH6 PRELIMINARY; PRT; 257 AA

DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
DE Related to SIS2 protein (cycle-specific gene control)

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;

RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Nemes H.W., Mannhaupt G.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RA German Neurospora genome project;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases  
 DR EMBL, AL451022; CAC18309.2; -.

DR Pfam; PF02441; Flavoprotein; 2.  
SQ SEQUENCE 257 AA; 28001 MW; B7744943B5998A6E CRC64;

Query Match	31.8%;	Score 364.5;	DB 3;	Length 257;
Best Local Similarity	38.9%;	Pred. No. 1.7e-28;		
Matches 84;	Conservative 31;	Mismatches 56;	Indels 45;	Gaps 7;

[illegible]

```
RESULT 8
09HC17 PRELIMINARY: PRT: 127 AA.
AC 09HC17:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MDS018.
GN MDS018.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang C., Qian B., Tu Y., Gu W., Wang Y., Han Z., Chen Z.:
RT "Novel genes expressed in hematopoietic stem/progenitor cells from
RT Myelodysplastic Syndromes patient."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF182419; AAG14955.1; -.
DR HSSP: Q9SWE5; 1E20.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
SQ SEQUENCE 127 AA; 14069 MW; 562D659DE77C0AF5 CRC64;

Query Match
Best Local Similarity 56.2%; Score 361; DB 4; Length 127;
Matches 68; Conservative 14; Mismatches 37; Indels 2; Gaps 1;

QY 89 MNKIGDPVLIHIELRRADVMIIAPLSANTLAKIAGLCDNLTCIVRAMDYSKPLFVAPA 148
DB 2 WKSRSPPVLIHIDLRWADLLVAPLADNTLGKVASGICDNLTCVARAMDSKPLFPCPA 61
QY 149 MNTLMNNEPTEHNLVLDLDELITLIPPIKKIACGDYNGMAEFSLIYSTR--LFW 206
DB 62 MNTAMEHBITAQOVDQLKAFGVEIPCAKRLVCGDEGAMAEVGIYDKKEVLFOH 121
QY 207 S 207
DB 122 S 122

RESULT 9
008438 PRELIMINARY: PRT: 674 AA.
AC 008438: 000019;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ORF_YOR054C.
GN YOR054C OR YOR529-05.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE OF 673-674 FROM N.A.
RA Landt O., Hiesel R., Unselid M.:
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Bohm C., Bolotin-Fukuhara M., Daignan-Fornier B., Dang D.V.,
RA Valens M.:
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS:
RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=97279235; PubMed=9133743;
RA Valens M., Bohm C., Daignan-Fornier B., Dang V., Bolotin-Fukuhara M.;
```

```
RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals the
RT presence of two tRNAs and 24 new open reading frames."
RL Yeast 13:379-390(1997).
DR EMBL: Z74961; CAA99246.1; -.
DR EMBL: Z70678; CAA94539.1; -.
DR HSSP: Q9SWE5; 1E20.
DR SGD: S0005580; YOR054C.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
SQ SEQUENCE 674 AA; 73648 MW; A97623473C8B6F605 CRC64;

Query Match
Best Local Similarity 36.6%; Score 323; DB 3; Length 674;
Matches 71; Conservative 40; Mismatches 57; Indels 26; Gaps 5;

QY 11 NKIEKKMNEVDVTKRKPRIILAAAGSVASIKFSNCHCFSEMAEYKAVASKSSLNFVOK 70
DB 395 NKQEEKQMA--STGEPSTL-----GGSRTYSNNSNVSOHPQLE----- 433
QY 71 PSLPQNVTLTYTDEDEWSSNKGDPVLIHIELRRADVMIIAPLSANTLAKIAGLCDNL 130
DB 434 --LPAHIQFWTDDDEMDVQRQRTDPLVLIHIELRRADILVAPLTANTLAKIALGCDNL 491
QY 131 TCIVRAMDYSKPLFVAPAMNTLMNNEPTEHNLVLDL--LGITLIPPIKKIAC--GDYG 187
DB 492 TSVIRAMNPFPIFLAPSMGSGTFNSIMTKHFRITIOEEMPVTVKPSKVGINGDIG 551
QY 188 NGMAEFSLIYSTRV 201
DB 552 LSGMDANETVGRKI 565

RESULT 10
066997 PRELIMINARY: PRT: 388 AA.
AC 066997:
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Pantothenate metabolism flavoprotein.
GN DFP OR AQ.815.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxId=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aufay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.:
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
RL EMBL: AE000708; AAC06944.1; -.
DR HSSP: Q9SWE5; 1E20.
DR InterPro: IPR005252; Cons.hypoth521.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
DR TIGRPFMS: TIGR00521; dfp; 1.
KW Complete proteome.
SQ SEQUENCE 388 AA; 43539 MW; 682BA44B64524D0B CRC64;

Query Match
Best Local Similarity 41.2%; Score 280.5; DB 16; Length 388;
Matches 70; Conservative 26; Mismatches 55; Indels 19; Gaps 7;

QY 30 ILLAAGSVASIKFSNCHCFSEMA---EVKAVASKSSLNFVOKPSLPONT---LYTD 82
DB 3 ILIGVCGGIASYSK---VCELVRELKRGHSVKITILPPAKFEM-SPLTROTISGNKAYTD 58
QY 83 EDWSSMNKIGDPVLIHIELRRADVMIIAPLSANTLAKIAGLCDNLTCIVRAMDYSKPL 142
```



Best Local Similarity 36.0%; Pred. No. 1.3e-18;  
Matches 64; Conservative 39; Mismatches 64; Indels 11; Gaps 4;

OY 30 ILLAASGSVASIKFSNLCHCFSEMA-EVKAVASKSSLNPFVDRKSP---LPON-VTLVTDDED 84  
DB 4 ILLAATGSIATSKASLVSSLKQGHQVYVLMQAAETFIQPLTLVYSONPVHLDVME 63  
OY 85 EWSNNKIDPVLHIELRRWADVMIIAPLSANTLAKIAGSLCDNLCTIVRAMDYSKPLF 144  
DB 64 PVP-----DQVNHIELGKKADLFIVVPATANTIAKLHGFADNMVTSIALALPSHPKL 117  
OY 145 VARAANTLMNNPETERHVLVLDDELGITLPIPKKLAGCDYNGAMAEPSLIYTVR 202  
DB 118 IAPAMNTKMYDHEVTQNNLTETTYGQILAPKESLACGDHGRGALADLTILLERIK 175

## RESULT 14

OY9X3X4 PRELIMINARY; PRT: 417 AA.

AC O9X3X4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
GN PanOtoluenate metabolism flavoprotein.  
OS Zymomonas mobilis.  
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
OC Zymomonas.  
OX NCBI\_TaxID=542;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ZM4;  
RA Lee H.D., Kang H.S.;  
RL "Sequence analysis of 42C11 fosmid clone of Zymomonas mobilis ZM4.";  
EMBL: AF088896; AAD21550.1; -;  
DR HSSP; Q9SWES; 1E20.  
DR InterPro; IPR005252; Cons\_hypoth521.  
DR InterPro; IPR003382; Flavoprotein.  
DR Pfam; PF02441; Flavoprotein; 1.  
DR TIGRFAMS; TIGR00521; dfp; 1.  
SQ SEQUENCE 417 AA; 45382 MW; 492P6585E5379138 CRC64;

Query Match 22.8%; Score 261; DB 2; Length 417;  
Best Local Similarity 35.2%; Pred. No. 8.2e-18;  
Matches 64; Conservative 39; Mismatches 73; Indels 6; Gaps 4;

OY 29 RILLAASGSVASIKFSNLCHCF-SEMAEYKAVASKSSLNPFVDRKSPQNVTLVTDDEWS 87  
DB 9 RILLVSGSIAIAIKAPDITIRLFRRKKADIRCLITTKGANFTPLALASLSGNFVADMD 68  
OY 88 SNKKIGDPVLHIELRRWADVMIIAPLSANTLAKIAGSLCDNLCTIVRAMDYSKPLFVAP 147  
DB 69 ESEAS--TRHRLAREAMIIIVAPASADFISIMAGLANDLASTVLAAD--SPLIVAP 124  
OY 148 AMNTLMNNPETERHVLVLDDELGITLPIPKKLAGCDYNGAMAEPSLIYTVRLEWE 206  
DB 125 ANNHRRMHHHSATQRNIHQKSDGISFVDEPAGAMACGETIGRLAPEDITLLSAESLFAE 184  
OY 207 SQ 208  
DB 185 EQ 186

## RESULT 15

OY9A254 PRELIMINARY; PRT: 412 AA.

AC O9A254;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE DNA/panOtoluenate metabolism flavoprotein.  
GN CC3712.

OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.

OY NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE006029; AAK25674.1; -;  
DR HSSP; Q9SWES; 1E20.  
DR TIGR; CC3712; -;  
DR InterPro; IPR005252; Cons\_hypoth521.  
DR InterPro; IPR003382; Flavoprotein.  
DR Pfam; PF02441; Flavoprotein; 1.  
DR TIGRFAMS; TIGR00521; dfp; 1.  
KW Complete proteome.  
SQ SEQUENCE 412 AA; 43105 MW; 9D4D24ED3A218515 CRC64;

Query Match 22.4%; Score 256.5; DB 16; Length 412;  
Best Local Similarity 32.1%; Pred. No. 2.3e-17;  
Matches 63; Conservative 39; Mismatches 73; Indels 21; Gaps 4;

OY 29 RILLAASGSVASIKFSNLCHCFSE-MAEYKAVASKSSLNPFVDRKSP-----QNVTL 79  
DB 19 RVLIIYGGGAAYKALLTLRLKAGVAVRPILTKGAAFPVPLSLAALAEKVYEDLS 78  
OY 80 YVDEDESSNNKIGDPVLHIELRRWADVMIIAPLSANTLAKIAGSLCDNLCTIVRAMDY 139  
DB 79 LTDEHEMG-----HIELSRADLVVAVPATADLAKAAGLAGDIASTTLATD- 127  
OY 140 SKPLFVAPAMNTLMNNPETERHVLVLDDELGITLPIPKKLAGCDYNGAMAEPSLIYS 199  
DB 128 -KPVLMAPAMNVMWLPATQRNIATLEADGVRFVPEEGAMACGEGGRLAEPALFA 186  
OY 200 TVRLFWEQARKOROG 215  
DB 187 AIMAALGPRAPLES 202

Search completed: November 25, 2002, 10:15:33  
Job time : 35 secs

115-09-868-300-8

Scoring table: BLOSUM62, Gapop 10.0, Gapext 0.5

Total number of hits satisfying c

Minimum	DB seq	length:	2000000000
Maximum	DB seq	length:	

Minimum Match	0%
Maximum Match	100%
Maximum first	45 summaries

SwissProt 40%

pred. No. is the number of cases in the score or criterion distribution score greater than or equal to the total score distribution and is derived by analysis of

## SUMMARIES

Result No.	Score	Query Length	DB	ID	Description
1	1060	92.6	201	1	HL3B_ARAHH
2	946	82.6	209	1	YL18_ARAHH
3	347	30.3	571	1	YL18_YEAST
4	331.5	29.0	562	1	SIS2_YEAST
5	328.5	28.3	531	1	SIS2_CNTR
6	239.5	20.9	402	1	DEP_SYN3
7	236.5	20.7	400	1	DEP_HAEN
8	217.5	19.0	455	1	DEP_STRMU
9	205.5	17.9	406	1	DEP_ECOLI
10	204.5	16.1	390	1	EPID_STAPB
11	184	15.4	418	1	DEP_BORBU
12	176.5	15.4	403	1	DEP_MCTU
13	135.5	11.8	403	1	DEP_METJA
14	109	9.5	184	1	PAAD_MNY3
15	108	9.1	182	1	PAAD_SYN3
16	104	9.0	209	1	PAAD_ARCFU
17	102.5	7.9	755	1	PAAD_PSEAE
18	90.5	7.8	181	1	PAAD_KTFAE
19	89.5	7.8	423	1	PAAD_PYRO
20	89.5	7.8	496	1	PAAD_YEAST
21	89	7.8	204	1	YH3_RICPR
22	88	7.7	186	1	PAAD_BACSU
23	87.5	7.6	186	1	PAAD_METJA
24	85	7.4	329	1	Y730_METJA
25	84	7.3	428	1	GSA_STAMM
26	83.5	7.3	192	1	GSA_STAMM
27	83	7.2	192	1	YE29_HELPI
28	82	7.2	329	1	PAAD_CHLNU
29	81	7.1	329	1	PAAD_CHLNU
30	80.5	7.0	187	1	YE29_HELPI
31	80	7.0	187	1	TKT_MYGE
32	80	7.0	211	1	PAAD_CAMDE
33	80	7.0	428	1	PAAD_VIBCH
					GSA_STAUU

[illegible]

## ALIGNMENTS

	HL3B-ARATH	STANDARD:	PRT: 201 AA.
HL3B-ARATH	RESULT 1		
ID	HL3B-ARATH	STANDARD:	
AC	P94063; Q9P62;	(rel. 40, Created)	
RC	16-OCT-2001	(last sequence update)	
DT	16-OCT-2001	(rel. 40, last annotation update)	
DT	16-OCT-2001	(rel. 40, last annotation update)	
DL	Haltolerance protein Hal13p.		
DN	HAL13 OR AT1G48610 OR RLN15.24.		
GC	Hal13p from Arabidopsis thaliana (mouse-ear cross).		
OS	Arabidopsis thaliana (Mouse-ear cross).		
OC	Eukaryota; Viridiplantae; Streptophyta; core eudicots; Rosidae;		
CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Arabidopsi		
CC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OK	MCL_TaxId=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND INDUCTION.		
RC	STRAN-ccv, Columbia;		
RC	MEDLIN=20117504; PubMed=10652125;		
RA	Aspinosa-Ruiz A., Belles J.M., Serrano R., Cullanez-Macia F.A.;		
RA	Arabidopsis thaliana plant growth.		
RL	Osmotic tolerance and plant growth.		
RL	Plant J. 20:529-539(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAN-ccv, Columbia;		
RC	MEDLIN=1016719; PubMed=11130712;		
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,		
RA	White O., Alonso J., Chao Q., Chen H., Cheuk R.F., Chin C.W., Dewar K.,		
RA	Buehler K., Chan L., Conway A.B., Conway A.R., Creasy T.H., Cui Y.,		
RA	Chung P., Burgu P., Feldblum T.V., Feng J.-D., Feng B., Fujii C.Y.,		
RA	Dunn P., Erdgu P., Feldblum T.V., Haas B., Hansen N.F., Hughes B., Huizar L.,		
RA	Gill J.E., Jenkins J., Johnson-Hopson C.C., Khan S., Khaykin E.,		
RA	Hunter J.L., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,		
RA	Kim C.J., Lee A., Lee J.M., Lewis J.S., Maiti R., Marshall A.,		
RA	Lam C.J., Liu X., Liu Z.A., Nguyen M., Niemann W.C., Osborne D.,		
RA	Lin X., Liu S.X., Liu Z.A., Nguyen M., Niemann W.C., Osborne D.,		
RA	Miltschcher J., Miranda M., Neufuss R., Rizzo M., Rooney T., Southwick A.M.,		
RA	Pai G., Peterson J.P., Pham P.K., Rizzo M., Rooney T., Southwick A.M.,		
RA	Sakano H., Tallon Van Aken S., Vayenberg M., Vysotskaya V.S., Walker M.,		
RA	Sun H., Tallon Van Aken S., Vayenberg M., Vysotskaya V.S., Walker M.,		
RA	Utterback T., Fraser C.M., Venner J.C., Davis R.W.;		
RA	"Sequence and analysis of chromosome 1 of the plant Arabidopsis		
RT	thaliana."		
RT	Nature 408:816-820(2000).		
RU	- FUNCTION INVOLVED IN PLANT GROWTH AND SALT AND OSMOTIC TOLERANCE.		
CC	Nature 408:816-820(2000).		
CC	- FUNCTION INVOLVED IN PLANT GROWTH AND SALT AND OSMOTIC TOLERANCE.		
CC	MAY CATALYZE THE ALPHA,BETA-DEHYDROGENATION OF A PEPTIDYL CITRINE		
CC	OF A CELLULOSE PARTNER (BY SIMILARITY).		
CC	- CORRELATOR: FMN (BY SIMILARITY).		
CC	- SUBUNIT: HOMODIMER (BY SIMILARITY).		
CC	- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, SHOOTS, LEAVES, FLOWERS,		
CC	DEVELOPING SILICLES AND SEEDS.		
CC	- INDUCTION: BY SALT STRESS.		
CC	- SIMILARITY: SOME TO YEAST AND C.TROPICALIS SIS7/NAL3 AND YEAST		
CC	YKL088W.		

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U80192; AAB53106.1; -  
DR EMBL: AC020889; AAF79709.1; ALT\_INT.  
DR HSP: Q9SWES; 1E20.  
DR InterPro: IPR003382; Flavoprotein.  
DR Pfam: PF02441; Flavoprotein; 1.  
KW Flavoprotein; FMN.  
SQ SEQUENCE 201 AA; 22415 MW; 7962655E408FA64C CRC64;

Query Match 92.6%; Score 1060; DB 1; Length 201;  
Best Local Similarity 100.0%; Pred. No. 2.7e-96;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MNNEVDVTYRKPRILLAAAGSVASIKFSNLCHEFSEMAEKAVAKSSLNFDKPSLPON 76  
DB 1 MNNEVDVTYRKPRILLAAAGSVASIKFSNLCHEFSEMAEKAVAKSSLNFDKPSLPON 60  
QY 77 VLYTDEDESSWNKIGDPVLIETLRNADVMTIAPLSANTLAKINGGLCDNLTCIVRA 136  
DB 61 VLYTDEDESSWNKIGDPVLIETLRNADVMTIAPLSANTLAKINGGLCDNLTCIVRA 120  
QY 137 WDYSKPLFVAPAMNTLMNNPFEERHLVLDLGTLLIPPIKKKLACG DYGNAMAEP SL 196  
DB 121 WDYSKPLFVAPAMNTLMNNPFEERHLVLDLGTLLIPPIKKKLACG DYGNAMAEP SL 180  
QY 197 IYTVRLFWESQARKORDGTS 217  
DB 181 IYTVRLFWESQARKORDGTS 201

#### RESULT 2

HL3A\_ARATH STANDARD; PRT; 209 AA.

AC Q9SWES; 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Halotolerance protein Hal3A.2.  
GN HAL3A OR ARG318030 OR MBG14.2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND INDUCTION.  
RC STRAIN=cv. Columbia;  
RA MEDLINE=20117504; PubMed=10652125;  
RT Espinosa-Ruiz A., Belles J.M., Serrano R., Cullanez-Macia F.A.;  
RT "Arabidopsis thaliana AtHAL3: a flavoprotein related to salt and  
RT osmotic tolerance and plant growth.";  
RT Plant J. 20:529-539(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA MEDLINE=20363099; PubMed=10907853;  
RT Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,  
RT TAC and BAC clones.";  
RT DNA Res. 7:217-221(2000).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RA MEDLINE=20444410; PubMed=10984663;  
RA Albert A., Martinez-Ripoll M., Espinosa-Ruiz A., Yennush L.,  
RA Cullanez-Macia F.A., Serrano R.;

RT "The X-ray structure of the FMN-binding protein AtHAL3 provides the  
RT structural basis for the activity of a regulatory subunit involved in  
RT signal transduction.";  
RL Structure 8:961-969(2000).  
CC -1- FUNCTION: INVOLVED IN PLANT GROWTH AND SALT AND OSMOTIC TOLERANCE.  
CC MAY CATALYZE THE ALPHA,BETA-DEHYDROGENATION OF A PEPTIDYL CYSTEINE  
CC OF A CELLULAR PARTNER.  
CC -1- COFACTOR: FMN.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, SHOOTS, LEAVES, FLOWERS,  
CC DEVELOPING STILOUES AND SEEDS WITH HIGHEST EXPRESSION IN SEED  
CC EMERYOS AND PHLOEM.  
CC -1- INDUCTION: BY SALT STRESS.  
CC -1- SIMILARITY: SOME, TO YEAST AND C.TROPICALIS SIS2/HAL3 AND YEAST  
CC YKL088W.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF166262; AAD51616.1; -  
DR EMBL: AB026641; BAB01331.1; -  
DR PDB: 1E20; 11-SEP-00.  
DR InterPro: IPR003382; Flavoprotein.  
DR Pfam: PF02441; Flavoprotein; 1.  
KW Flavoprotein; FMN; 3D-structure.  
SQ SEQUENCE 209 AA; 23355 MW; 3AB1B8364F8E40DE CRC64;

Query Match 82.6%; Score 946; DB 1; Length 209;  
Best Local Similarity 85.6%; Pred. No. 3.7e-85;  
Matches 173; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 16 MNNEVDVTYRKPRILLAAAGSVASIKFSNLCHEFSEMAEKAVAKSSLNFDKPSLPQ 75  
DB 8 MNNEVDVTYRKPRILLAAAGSVASIKFSNLCHEFSEMAEKAVAKSSLNFDKPSLPQ 67  
QY 76 NVLYTDEDESSWNKIGDPVLIETLRNADVMTIAPLSANTLAKINGGLCDNLTCIVR 135  
DB 68 NVLYTDEDESSWNKIGDPVLIETLRNADVMTIAPLSANTLAKINGGLCDNLTCIVR 127  
QY 136 AMYDSCPFLFVAPAMNTLMNNPFEERHLVLDLGTLLIPPIKKKLACG DYGNAMAEP SL 195  
DB 128 AMYDSCPFLFVAPAMNTLMNNPFEERHLVLDLGTLLIPPIKKKLACG DYGNAMAEP SL 187  
QY 196 IYTVRLFWESQARKORDGTS 217  
DB 188 IYTVRLFWESQARKORDGTS 209

#### RESULT 3

YK18\_YEAST STANDARD; PRT; 571 AA.

AC P36076; 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 65.2 kDa protein in MIF2-CYT2 intergenic region.  
GN YKL088W.  
OS Saccharomyces cerevisiae (Baker's Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Pohl T.M., Pohl F.M.;  
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO YEAST AND C.TROPICALIS SIS2/HAL3. SOME, TO  
CC A.THALIANA HAL3A AND HAL3B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; 228088; CAA81926.1; -  
 DR PIR; S37913; S37913.  
 DR HSSP; Q9SWE5; IE20.  
 DR SGD; S0001571; YKL088W.  
 DR InterPro; IPR003382; Flavoprotein.  
 DR Pfam; PF02441; Flavoprotein; 1.  
 DR Hypothetical protein.  
 KM DOMAIN 508 570 ASP/GLU-RICH (HIGHLY ACIDIC).  
 SQ SEQUENCE 571 AA: 65238 MW; 96C74C2394EFC6AB CRC64;

Query Match 30.3%; Score 347; DB 1; Length 571;  
 Best Local Similarity 37.6%; Pred. No. 3.8e-26;  
 Matches 89; Conservative 39; Mismatches 77; Indels 32; Gaps 8;

OY 8 LGCNIEKKMNEVDVTRKPR-----LLAASGSVASIK---FSNIC 47  
 DB 274 LSANNSTNR-NKEVITAPTPGPRVPTFEQKEDDKKFKHLIGATSGVATIKVPLIIDKLF 332  
 OY 48 HCFSS-EMAEKVAASKSLNFEVKPSLPONTLYTDEDEM--SSNNKIG-----DPVLEHT 99  
 DB 333 KLYGEKISIQILVTRKPAHEFLKGLKSHVAKIMREEDMVDVAVKKNSTLSLNILH 392  
 OY 100 ELRRADVMIIPLSANTLAKIAGLCNDLTCIVRAMDYSKPLEVAPAMNTLMNNPFT 159  
 DB 393 ELRKWADIFLIPLSANTLAKLANGICNNLITSVMRDSPLPVLILAPAMNTFMYINPMT 452  
 OY 160 EKHVLV--DELGITLIPPIKIKKLAGCGNGAMAEPSLIYSTVRLFMFSQARKQD 214  
 DB 453 KKHLSLVODYPIFOVLKPEKVLICGDI GCMGRMTDIVEIVRR-RINEIRKARD 508

## RESULT 4

SIS2\_YEAST  
 ID SIS2\_YEAST STANDARD; PRT; 562 AA.  
 AC P36024.  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE SIS2 protein (Halotolerance protein HAL3).  
 GN SIS2 OR HAL3 OR YKR072C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95220693; PubMed=7705654;  
 RA di Como C.J., Bose R., Arndt K.T.;  
 RT "Overexpression of SIS2, which contains an extremely acidic region,  
 RT increases the expression of SMY4, CLN1 and CLN2 in sld4 mutants.";  
 RL Genetics 139:95-107(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Pohl T.M., Pohl F.M.;  
 RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP CHARACTERIZATION.  
 RA MEDLINE=96009574; PubMed=7565698;  
 RA Ferrando A., Krom S.J., Rios G., Fink G.R., Serrano R.;  
 RT "Regulation of cation transport in Saccharomyces cerevisiae by the  
 RT salt tolerance gene HAL3.";  
 RL Mol. Cell. Biol. 15:5470-5481(1995).  
 RN [4]  
 RP IDENTIFICATION AS INHIBITORY SUBUNIT OF PP21.  
 RA MEDLINE=96301578; PubMed=9636153;

RA de Nadal E., Clotet J., Posas F., Serrano R., Gomez N., Arino J.;  
 RT "The yeast halotolerance determinant Halp is an inhibitory subunit of  
 RT the Ppz1p Ser/Thr protein phosphatase.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:7357-7362(1998).  
 CC -I- FUNCTION: MAY STIMULATE EXPRESSION OF CERTAIN GENES THAT ARE  
 CC PERIODICALLY EXPRESSED DURING LATE G1. ALSO MODULATES THE  
 CC EXPRESSION OF THE ENA1 ATPASE. INTERACTS WITH THE C-TERMINAL  
 CC DOMAIN OF THE SERINE-THREONINE PROTEIN PHOSPHATASE PP21 AND ACTS  
 CC AS AN INHIBITORY SUBUNIT OF PP21.  
 CC -I- SUBCELLULAR LOCATION: NUCLEAR OR CYTOPLASMIC.  
 CC -I- SIMILARITY: TO C.TROPICALIS SIS2/HAL3 AND TO YEAST YKL088W. SOME,  
 CC TO A.THALIANA HAL3A AND HAL3B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; 001878; AAA80000.1; -  
 DR EMBL; 228297; CAA82151.1; -  
 DR PIR; S38149; S38149.  
 DR HSSP; Q9SWE5; IE20.  
 DR SGD; S0001780; SIS2.  
 DR InterPro; IPR003382; Flavoprotein.  
 DR Pfam; PF02441; Flavoprotein; 1.  
 DR Protein phosphatase inhibitor; Nuclear protein.  
 KM DOMAIN 496 553 ASP/GLU-RICH (HIGHLY ACIDIC).  
 FT SEQUENCE 562 AA: 62478 MW; 19A9A475145DA7AB CRC64;

Query Match 29.0%; Score 331.5; DB 1; Length 562;  
 Best Local Similarity 35.6%; Pred. No. 1.2e-24;  
 Matches 77; Conservative 41; Mismatches 67; Indels 31; Gaps 6;

OY 2 FETSSFLGCKIEKKMNEVDVTRKPRILLAAAGSVASIKFSNLCHEFSMAEYKAAVAS 61  
 DB 268 FGATGSLVFKI-KPMIKKLEIYGRDI-----SIQVILQSATQFPEQRYTKIKKS 320  
 OY 62 KSSLN----FVDRPS-----LPQNVITYTDEDESSNNKIGDPVLEHT 100  
 DB 321 SEKLKMSQYESTPATPVTPPGQCMAQVLELPPIQLMTDQDEDAKQRPDPLHLE 380  
 OY 101 LRRMADVMIIPLSANTLAKIAGLCNDLTCIVRAMDYSKPLEVAPAMNTLMNNPFT 160  
 DB 381 LRRMADILVAVPLTANTLSKIALGLCDNLITSYIRAMNSYFILLAPSVSSTFNSMTRK 440  
 OY 161 RHLVLDE--LGITLIPPIKRLAC-GDYNGAMAE 193  
 DB 441 KQLQTIKEMSWVTYFKPSEKVMINDIGLGGMMD 476

## RESULT 5

SIS2\_CANTR  
 ID SIS2\_CANTR STANDARD; PRT; 531 AA.  
 AC 012600.  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE SIS2 protein (Halotolerance protein HAL3).  
 GN SIS2 OR HAL3.  
 OS Candida tropicalis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5482;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=NCYC 2512;  
 RC MEDLINE=97082501; PubMed=8923737;  
 RA Rodriguez P.L., Ali R., Serrano R.;  
 RT "Ctdc5p and Cth13p: two putative regulatory proteins from Candida  
 RT tropicalis with long acidic domains.";

RL Yeast 12:1321-1329(1996).  
CC - FUNCTION: MAY STIMULATE EXPRESSION OF CERTAIN GENES THAT ARE  
CC PERIODICALLY EXPRESSED DURING LATE G1. ALSO MODULATES THE  
CC EXPRESSION OF THE ENAI ATPASE (BY SIMILARITY).  
CC - SUBCELLULAR LOCATION: NUCLEAR OR CYTOPLASMIC (BY SIMILARITY).  
CC - SIMILARITY: TO YEAST SIS2/HAL3 AND TO YEAST YKL088W. SOME, TO  
CC A THALIANA HAL3A AND HAL3B.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X88900; CA61362.1; -  
DR HSSP: Q9SWE5; IE20.  
DR InterPro: IPR003382; Flavoprotein.  
DR Pfam: PF02441; Flavoprotein; 1.  
FT DOMAIN 233 240 POLY-THR.  
FT DOMAIN 246 249 POLY-THR.  
FT DOMAIN 484 513 ASP/GLU-RICH (HIGHLY ACIDIC).  
SQ SEQUENCE 531 AA: 58171 MW; D5FF196BB957B14 CRC64;  
Query Match 28.3%; Score 324.5; DB 1; Length 531;  
Best Local Similarity 36.6%; Pred. No. 5.5e-24;  
Matches 78; Conservative 46; Mismatches 60; Indels 29; Gaps 9;  
OY 4 TSSPLQCKNIEKKMNEVTVTRKPR-----TLAASG--SVASIKF--SNLCGCF-SE 52  
DB 240 TNSNSNTTTTKEQNSNID--PRLPODDGKPHVLIGVCGALSGKVKLVNKLLEYTSD 297  
OY 53 MAEVAKAVASKSLNFWKDRSLPO-----NVTLYTDEDESSNKKIGDPYLHIELRR 103  
DB 298 KTSIOVILTKSENEL-----LPETLVNLENVKKVRYWTDIDETTKTRLDPLHLIELRR 353  
OY 104 WADVMIAPLSANTLAKIAGLCNLLTCIVRAMDYSKPLFVAPAMNTLMNPFTEHRL 163  
DB 354 WADILLVCLPTANTLAKISLGICDNLNVIRAMNSYPIILLAPADSHSYSSSTTKRDL 413  
OY 164 VLL--DELGITLIPPIKKRL-ACGDYGNCAMAE 193  
DB 414 RLADDMPIEVLKPLEKVFSGYDGMGMD 446  
RESULT 6  
DFF\_SYNY3  
ID DFF\_SYNY3 STANDARD: PRT; 402 AA.  
AC P73881;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA/pantothenate metabolism Flavoprotein homolog.  
GN DFP OR SL0250.  
OS Synecchocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecchocystis.  
OX NCBI\_Taxid=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki C., Nakazaki N., Naruo K.,  
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Matanabe A.,  
RA Yamada M., Yasuda M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synecchocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-116(1996).  
CC - FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE  
CC METABOLISM (BY SIMILARITY).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: D90910; BAA17944.1; -  
DR HSSP: Q9SWE5; IE20.  
DR InterPro: IPR003382; Cons\_hypoth521.  
DR InterPro: IPR003382; Flavoprotein.  
DR Pfam: PF02441; Flavoprotein; 1.  
DR TIGRFAMs: TIGR00521; dfp; 1.  
KW Flavoprotein; Complete proteome.  
SQ SEQUENCE 402 AA: 43419 MW; 7E3F73DE185EC97C CRC64;  
Query Match 20.9%; Score 239.5; DB 1; Length 402;  
Best Local Similarity 37.3%; Pred. No. 7.7e-16;  
Matches 66; Conservative 25; Mismatches 69; Indels 17; Gaps 6;  
OY 29 RLLAASGSVASIKSEN-LCHCFSEMAEYKAVAKSKSLNFWKPSLPQWNT-----LYT 81  
DB 6 RLILGVGGIAAKICEVVSQLEFQGAEVRYITAEKRV-----TPLLFTTLARPAYG 61  
OY 82 DEDEWSSNKKIGDPYLHIELRWADVIITAPLSANTLAKIAGLCNLLTCIVRAMDYSK 141  
DB 62 DAD---FWQPIHHRPLHLDIGEMADIFLAPLHATLAKLGHFADLLSNVLA--SSG 116  
OY 142 PLFVAPAMNTLMNPFTEHRL-VLDELGITLIPPIKKRLACGDYGNCAMAEPSLI 197  
DB 117 PLLAPAMNTDMWEDQAVQNRNLQQLIDRRYHLLAPNGLLACDRGVRLAEPAOI 173  
RESULT 7  
DFF\_HAEIN  
ID DFF\_HAEIN STANDARD: PRT; 400 AA.  
AC P44953;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE DNA/pantothenate metabolism Flavoprotein homolog.  
GN DFP OR HI0953.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
CC Haemophilus.  
OX NCBI\_Taxid=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-Rd / KW20 / ATCC 51907;  
RC MEDLINE=9530630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Ullnerbach T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd.";  
RL Science 269:496-512(1995).  
CC - FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE  
CC METABOLISM (BY SIMILARITY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----



```

CC -----
DR EMBL: U32776; AAC22614.1; -
DR HSSP: Q9SWE5; IE20.
DR TIGR: H10953; -
DR InterPro: IPR005252; Cons_hypoth521.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein.1.
DR TIGRfams: TIGR00521; dfp.1.
DR Flavoprotein: Complete proteome.
KW SEQUENCE 400 AA; 43401 MW; 1D6FEAE20C26B913 CRC64;

Query Match 20.7%; Score 236.5; DB 1; Length 400;
Best Local Similarity 32.3%; Pred. No. 1.5e-15;
Matches 61; Conservative 38; Mismatches 67; Indels 23; Gaps 4;

QY 30 ILLAASGSVASIKFNSLCHCFSE-MAEVKAVAKSSSLNFVDKPSLPQNTLYTDEDEMS 88
DB 8 IYVGITGGIAAKYKTELRLKKAEEVAVVLTTPAAAEVTPTLT-----QALS 56
QY 89 WNKIGDPVL-----HIELRMADVMIAPLSANTLAKIAGLCNLTLCYIRAMDY 139
DB 57 GNAVSQSLDDPOAELAMGHIELAKWADAIILAPASADFIARLTIGMANDLSTICLA--T 114
QY 140 SKPLVAPAPAMNTLMNNPTEERHVLDELGITLPIPKKACGPGYGNAAEPLIYS 199
DB 115 NAPIFLAPAMNQMYHQSITQONLTTLQTRGIELIGPNSGFQACGDMGKRMSEPEIFT 174
QY 200 TVRLFWESQ 208
DB 175 ALSDFFSQK 183

RESULT 8
DFP_STRMU STANDARD; PRT: 145 AA.
AC 054433:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein homolog (Fragment).
GN DFP.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG8.
RX MEDLINE=96312359; PubMed=8763945;
RA Gutierrez J.A., Crowley P.J., Brown D.P., Hillman J.D.,
RA Youngman P., Bleiweis A.S.;
RT "Insertional mutagenesis and recovery of interrupted genes of
RT Streptococcus mutans by using transposon Tn917: preliminary
RT characterization of mutants displaying acid sensitivity and
RT nutritional requirements."
RL J. Bacteriol. 178:4166-4175(1996).
CC -!- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U48885; AAC44502.1; -
CC HSSP: Q9SWE5; IE20.
CC InterPro: IPR003382; Flavoprotein.
CC Pfam: PF02441; Flavoprotein.1.
KW Flavoprotein.1.
FT NON_TER 1 1

```

```

FT NON_TER 145 145
SQ SEQUENCE 145 AA; 15819 MW; 5920CDD57C8F2DAB CRC64;

Query Match 19.0%; Score 217.5; DB 1; Length 145;
Best Local Similarity 35.2%; Pred. No. 3e-14;
Matches 51; Conservative 30; Mismatches 57; Indels 7; Gaps 3;

QY 29 ILLAASGSVASIKFNSLCHCFSEMA-EVKAVAKSSSLNFVDKPSLPQNTLYTDEDEMS 87
DB 4 KILLAVSGSIAAYKADLSHQLTGLGYHNVFMTNNAKQFIP----PLTLYVLSKRPVYS 59
QY 88 SNKIGDP--VLHIELRMADVMIAPLSANTLAKIAGLCNLTLCYIRAMDYSPFLV 145
DB 60 NWKEDDPQVINIHALAKADFLPFPASANTLAHLHGFDADIVTSVALAPLEVPKFF 119
QY 146 APAMNTLMNNPTEERHVLDELG 170
DB 120 APAMNTKMTENPTQSNITLLKKFG 144

RESULT 9
DFP_ECOLI STANDARD; PRT: 406 AA.
AC P24285; P76718;
DT 01-MAR-1992 (Rel. 21, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein.
GN DFP OR B3639.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL2 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL Genomics 16:551-561(1993).
RN [2]
RP SEQUENCE OF 288-406 FROM N.A.
RC STRAIN=KL2;
RX MEDLINE=84057777; PubMed=6139280;
RA Lundberg L.G., Thoresson H.O., Karlstrom O.H., Nyman P.O.;
RT "Nucleotide sequence of the structural gene for dnfPase of
RT Escherichia coli K-12."
RL EMO J. 2:967-971(1983).
CC -!- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: F10328; AAA61992.1; ALT_INIT.
CC EMBL: AE000441; AAC76663.1; ALT_INIT.
CC EMBL: V01578; -; NOT_ANNOTATED_CDS.
CC HSSP: Q9SWE5; IE20.
CC Ecogene: E610004; dfp.
CC InterPro: IPR005252; Cons_hypoth521.
CC InterPro: IPR003382; Flavoprotein.
CC Pfam: PF02441; Flavoprotein.1.
CC TIGRfams: TIGR00521; dfp.1.
KW Flavoprotein: Complete proteome.
SQ SEQUENCE 406 AA; 43438 MW; CBD11B9347E8C6AB CRC64;

Query Match 17.9%; Score 205.5; DB 1; Length 406;

```





```
OY 62 KSLNLFVDPKPSLPQNVTLTYLDEDEWSSMKN1-----GDPVLHIELRRMADVMIIAPLSAN 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 EETKRIIGKEAL-----KFGCGNEVEYEITGD-IEHILLYNECCLLIYPATAN 105
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 117 TLAKIAGLCDNLITFCIVAMDYSKPLFVAPAMNTLMNNPFTERRHLVLLDELG-ITLIP 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 106 IISKINLGIADINVTALMFEGNKPIFTVPAMHENMFN--AIKRHIIDKLEKDKYIIS 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 176 P 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 164 P 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
PAAD_METJJA STANDARD: PRT: 184 AA.
AC 057566:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
GN MJ0102.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-JAI-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerelevage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overberg R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RT Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U67467; AAB98082.1; -
CC TIGR: MJ0102;
DR InterPro: IPR003382; Flavoprotein.
DR InterPro: IPR004507; Ubix.
DR Pfam: PF02441; Flavoprotein; 1.
DR TIGRFAMs: TIGR00421; ubix; 1.
DR Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 184 AA; 20643 MW; F3EDBF5B48A82C5 CRC64;

Query Match
Best Local Similarity 9.5%; Score 109; DB 1; Length 184;
Matches 42; Conservative 24; Mismatches 61; Indels 40; Gaps 7;
```

```
OY 138 DVS-----KPLFVAPAMNTLMNNPFTERRHL---VLLDELGITLIPPI 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 106 DIALERRKLIIMPR-----EMPFNSIHLENMLKLSNIGALVMPPI 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
PAAD_SYNY3 STANDARD: PRT: 206 AA.
AC P27243;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
GN SLR1099.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RT DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: D90900; BAA16758.1; -
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
DR Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 206 AA; 22172 MW; 7C1E13F47996E3F CRC64;

Query Match
Best Local Similarity 9.4%; Score 108; DB 1; Length 206;
Matches 48; Conservative 24; Mismatches 56; Indels 52; Gaps 10;

OY 30 ILLAAG-----SVASIKFSNLCHEFSEMAEYKAVASKSLNF-----VDKPSLPQNV 77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 LILGVSAGSLIYAVRAIK-----HLLADYITELVYASRSYVWQAEQNIQGPES-- 57
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 78 TLTYDEDEWSSW--SKIGDPVLHIELRRMADY-----MIAPLSANTLANI 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 ---AQAEFMSQAGVEKGLKCH-----RMGDVGATIASGSRGACGMVYLPCSMSTYAKL 110
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 122 AGGLCDNLITCIYRANDY-----SKPLFVAPAMNTLMNNPFTERRHLVLLDELGITLIPPI 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 111 AVGMSSDLLE---RAADVOIKEGKPLVVVPRETP--SLIHRLNLTSLAEAGRTVPAL 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: November 25, 2002, 10:14:53  
Job time : 13 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2002, 10:13:41 ; Search time 21 Seconds

(Without alignments)  
993.389 Million cell updates/sec

Title: US-09-868-300-8

Perfect score: 1145

Sequence: 1 EFGTSFLGCKIKKKMMME.....YSTVRLFWESQARKQDQTS 217

Scoring table: BLOSUM62

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR-7380

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	421.5	36.8	625	2 T37719	thymidylate syntha
2	421	36.8	237	2 T21376	hypothetical prote
3	347	30.3	571	2 S37913	hypothetical prote
4	331.5	29.0	562	2 S38149	S152 protein - yea
5	324.5	28.3	531	2 S57752	hal3 protein - yea
6	323	28.2	674	2 S66937	probable membrane
7	280.5	24.5	388	2 B70371	pantothenate metab
8	270.5	23.6	410	2 A82194	pantothenate metab
9	264.5	23.1	183	2 F98010	pantothenate metab
10	264.5	23.1	183	2 H95142	N-terminal region
11	256.5	22.4	412	2 F87709	Flavoprotein limpo
12	249.5	21.8	394	2 A81058	DNA/pantothenate m
13	246.5	21.5	394	2 DB1819	DNA/pantothenate m
14	243.5	21.3	178	2 DB6696	hypothetical prote
15	240	21.0	404	2 A13538	pantothenate metab
16	239.5	20.9	402	2 S75082	phosphopantocheny
17	239.5	20.9	404	2 F83963	pantothenate metab
18	237.5	20.7	398	2 C97112	Flavoprotein dfr f
19	236.5	20.7	400	2 G64104	Flavoprotein invol
20	232.5	20.3	399	2 F89893	pantothenate metab
21	228.5	20.0	399	2 A11674	hypothetical prote
22	227.5	19.9	399	2 A11302	pantothenate metab
23	226.5	19.8	405	2 F81354	pantothenate metab
24	225	19.7	405	2 A10006	DNA/pantothenate
25	224.5	19.6	406	2 D69878	DNA/pantothenate m
26	224	19.6	425	2 E71889	pantothenate metab
27	223.5	19.5	402	2 B82962	pantothenate metab
28	221	19.3	425	2 A64625	DNA/pantothenate m
29	216.5	18.9	399	2 A82351	DNA/pantothenate m

30	210.5	18.4	394	2 A72223	pantothenate metab
31	210.5	18.4	409	2 H82842	DNA/pantothenate m
32	207.5	18.1	407	2 AC0971	conserved hypotnet
33	205.5	17.9	430	1 A65165	pantothenate metab
34	205.5	17.9	430	2 B91193	pantothenate metab
35	205.5	17.9	430	2 C86040	pantothenate metab
36	204.5	17.9	181	2 S23418	pantothenate metab
37	204	17.8	401	2 AD2615	lactidolic epider
38	184	16.1	390	2 C70201	pantothenate metab
39	183.5	16.0	185	2 B95019	Flavoprotein limpo
40	176.5	15.4	367	2 C97397	pantothenate metab
41	176.5	15.4	418	2 E70899	pantothenate metab
42	173.5	15.2	419	2 G66976	probable Flavoprot
43	169	14.8	401	2 E75114	dna/pantothenate m
44	168.5	14.7	413	2 B90162	DNA/pantothenate m
45	167.5	14.6	401	2 G71018	pantothenate metab

ALIGNMENTS

```
RESULT 1
T37719
thymidylate synthase-like DNA metabolism protein - fission yeast (Schizosaccharomyces
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37719
R:Murphy, L., Harris, D., Barrell, B.G., Rajandream, M.A., Lyne, M.H.
Submitted to the EMBL Data Library, August 1999
A:Reference number: Z21739
A:Accession: T37719
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-625 <MUP>
A:Cross-references: EMBL:AL109770; PIDN:CB52423.1; GSPDB:GN00066; SPDB:SPAC15E1.04
A:Experimental source: strain 972h-; cosmid c15E1
C:Genetics:
A:Gene: SPDB:SPAC15E1.04
A:Map position: 1

Query Match      36.8%; Score 421.5; DB 2; Length 625;
Best Local Similarity 48.8%; Pred. No. 6.1e-33;
Matches 84; Conservative 31; Mismatches 52; Indels 5; Gaps 3;

QY 27 KPRILLASGVASIKFSNLCHCFSEW--AEYKAVASKSLFVDPKPSLPQ-NNTLYTDE 83
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 30 KHIILVAATGSVAATKILTVKSLTYGVDVQVLTDPARNFVEKEDITLGVVYNNNA 89
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 84 DEMSSMNRIGDPVLEHIEELRRADVMIIAPLSANTLAKTAGICDNLTCIYRWDYSKPL 143
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 90 DDMKMWMDGIECPITHEIELRRMHLLILAPLSANTMAKMANGCODLTLISLRAMPPLKPI 149
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 144 EVAPAMNTLMNNPFTERRHVLDEL--GTLIPPIKKLACGDYGAMAE 193
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 150 ILAPAMNTLMNTNPITOEHLAISRIYKNSFIMPIERVLACGDIGMGMAE 201
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
T21376
hypothetical protein F25H9.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21376
R:Smey, R.
Submitted to the EMBL Data Library, October 1996
A:Reference number: Z19414
A:Accession: T21376
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-237 <WII>
A:Cross-references: EMBL:Z81069; PIDN:CA802993.1; GSPDB:GN00023; CESP:F25H9.6
A:Experimental source: clone F25H9
C:Genetics:
```



OY 164 VLL--DELGITLPIPIKKKL-ACGDYNGAMAE 193  
DB 414 RLADMPWIEVLKPLEKVGSTGIDIGMGMTD 446

## RESULT 6

S66937

probable membrane protein YOR054c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein O2801

C:Species: *Saccharomyces cerevisiae*

C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002

C:Accession: S66937; S66928

R:Bohr, C.; Bolotin-Fukuhara, M.; Baigian-Fornier, B.; Dang, D.V.; Valens, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66929

A:Accession: S66937

A:Molecule type: DNA

A:Residues: 1-674 &lt;BOH&gt;

A:Cross-references: EMBL:Z74961; NID:g1420188; PID:e251974; PID:g1420190; MIPS:YOR054c

A:Experimental source: strain S288C

R:Landt, O.; Hiesel, R.; Unseld, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66907

A:Accession: S66928

A:Molecule type: DNA

A:Residues: 674 &lt;LAN&gt;

A:Cross-references: EMBL:Z74961; MIPS:YOR054c

A:Experimental source: strain S288C

C:Genetics:

A:Cross-references: SGD:S0005580

A:Map position: 15R

C:Keywords: Transmembrane protein

F:468-484/Domain: transmembrane #status predicted &lt;TM&gt;

Query Match 28.2%; Score 323; DB 2; Length 674;

Best Local Similarity 36.6%; Pred. No. 2.7e-23;

Matches 71; Conservative 40; Mismatches 57; Indels 26; Gaps 5;

OY 11 NIEKKMNEVDVTRKPRLLAASGSVASIKFSNLCHGSEMAEYKAAVASKSLNFVDR 70  
DB 395 NQOEKQNA--STTGFPSTL-----GGSRTYSNSNVVSGHPQIE----- 433

OY 71 PSLPQNVLTYTEDEKSSNNKIGDPVLIHELRRMADVMITAPLSANTLAKIAGLCDNLL 130  
DB 434 --LPAIDIQMTODEMDVWRQRTDPLVLIHELRRMADILVAPLTANTLAKIAGLCDNLL 491

OY 131 TCIVRAMDPSKPLFVAPAMNTLMNMPFERHLVILDE--LGTLLPIPIKKKLAC-GDG 187  
DB 492 TSVIRAMNPTPLFLAPSKSGCTFNSIMTKHFRITIQEDMPVTVKPSKVGWINGDIG 551

OY 188 NGAMAEPLIYSTV 201  
DB 552 LSGMDANEIVGKI 565

OY 188 NGAMAEPLIYSTV 201  
DB 552 LSGMDANEIVGKI 565

## RESULT 7

B70371

pantothenate metabolism flavoprotein dfp - *Aquifex aeolicus*

N:Alternate names: probable aspartate 1-decarboxylase activase

C:Species: *Aquifex aeolicus*

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 29-Sep-1999

C:Accession: B70371

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: B70371

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-388 &lt;AO&gt;

A:Cross-references: GB:AE000708; NID:g2983356; PIDN:AAC06944.1; PID:g2983357; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:  
A:Gene: dfp  
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 24.5%; Score 280.5; DB 2; Length 388;

Best Local Similarity 41.2%; Pred. No. 1.9e-19;

Matches 70; Conservative 26; Mismatches 55; Indels 19; Gaps 7;

OY 30 ILLAASGSVASIKFSNLCHGSEMA---EYKAAVASKSLNFVDRKPSLPQNT--LYTD 82  
DB 3 ILIGVGGIASYK---VCELVRKLKRGHSVKITLPPFAKEFM-SPLTPQTLISGNKAYTD 58

OY 83 EDEWSSWNKIGDPVLIHELRRMADVMITAPLSANTLAKIAGLCDNLLTCIVRAMDYSKP 142  
DB 59 KD---WEE--EPLAHINLARWADVLIAPAPAMNTLAKIANGISDNLTTTILA--YGRP 110

OY 143 LPVAPAMNTLMNMPFERHLVILDELTGTLPIPIKKKLACGDYNGAMA 192  
DB 111 LTVAPAMNTVMKSPSTQENLKLKEMGHVILPEFGVLACKVEGKILA 160

OY 143 LPVAPAMNTLMNMPFERHLVILDELTGTLPIPIKKKLACGDYNGAMA 192  
DB 111 LTVAPAMNTVMKSPSTQENLKLKEMGHVILPEFGVLACKVEGKILA 160

## RESULT 8

AH2194

pantothenate metabolism flavoprotein [imported] - *Nostoc* sp. (strain PCC 7120)C:Species: *Nostoc* sp.A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AH2194

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Matsumoto, A.; Iriku

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata

DNA Res. 8, 205-213, 2001

A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AH2194

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-410 &lt;KUR&gt;

A:Cross-references: GB:BA000019; PIDN:BA074810.1; PID:g17132205; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all111

Query Match 23.6%; Score 270.5; DB 2; Length 410;

Best Local Similarity 38.0%; Pred. No. 1.9e-18;

Matches 76; Conservative 29; Mismatches 74; Indels 21; Gaps 7;

OY 26 KRPRILLAASGSVASIK---FSNLCHGSEMAEYKAAVASKSLNFVDRKPSLPQNTL-- 79  
DB 11 RKSRYLVAVGGGIAAYKCEVVSNTL---FKSGVEYVILTRSAOEFITPLTSL--TLSR 64

OY 80 ---YTPDEKSSWNKIGDPVLIHELRRMADVMITAPLSANTLAKIAGLCDNLLTCIVRA 136  
DB 65 HPAITDDDFWQSTH---SRPLHIEGEMADLVIAPLTANTLAKIAGLCDNLLTITVLA 121

OY 137 WDYSKPLFVAPAMNTLMNMPFERH-LVLDDELGTLLPIPIKKKLACGDYNGAMAEPS 195  
DB 122 --SNCFALLAPAMNTDMWQGLTVQRWQQLTDSRYHGIGTASGLLACDRVAGARRAEP 179

OY 196 LIYSIVRLFWESQARKQKRG 215  
DB 180 EILAYTQSLHHTQGRKDLAG 199

OY 137 WDYSKPLFVAPAMNTLMNMPFERH-LVLDDELGTLLPIPIKKKLACGDYNGAMAEPS 195  
DB 122 --SNCFALLAPAMNTDMWQGLTVQRWQQLTDSRYHGIGTASGLLACDRVAGARRAEP 179

OY 196 LIYSIVRLFWESQARKQKRG 215  
DB 180 EILAYTQSLHHTQGRKDLAG 199

## RESULT 9

F98010

N-terminal region of Dfp protein homolog [imported] - *Streptococcus pneumoniae* (strainC:Species: *Streptococcus pneumoniae*

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 03-Jun-2002

C:Accession: F98010

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.

y, P.; Sun, P.M.; Winkler, M.E.







**THIS PAGE BLANK (USPTO)**







Db 400 GGAGCVCILLMDVVSADNYTLDLW--AGQQLWKGSF 435

## RESULT 9

```

US-09-935-868-38
? Sequence 38, Application US/09935868
? Patent No. US20020164690A1
? GENERAL INFORMATION:
? APPLICANT: Regeneron Pharmaceuticals, Inc
? TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using
? FILE REFERENCE: REG 203D
? CURRENT APPLICATION NUMBER: US/09/935,868
? CURRENT FILING DATE: 2002-04-11
? PRIOR APPLICATION NUMBER: PCT/US99/22045
? PRIOR FILING DATE: 1999-09-22
? NUMBER OF SEQ ID NOS: 52
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 38
? LENGTH: 780
? TYPE: PRT
? ORGANISM: Homo Sapiens
? US-09-935-868-38

```

## RESULT 10

```

US-09-935-868-42
: Sequence 42, Application US/09935868
: Patent No. US20020164690A1
: GENERAL INFORMATION:
: APPLICANT: Regeneron Pharmaceuticals, Inc
: TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using
: FILE REFERENCE: REG 203D
: CURRENT APPLICATION NUMBER: US/09/935,868
: CURRENT FILING DATE: 2002-04-11
: PRIOR APPLICATION NUMBER: PCT/US99/22045
: PRIOR FILING DATE: 1999-09-22
: NUMBER OF SEQ ID NOS: 52
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 42
: LENGTH: 780
: TYPE: PRT
: ORGANISM: Homo Sapiens
: US-09-935-868-42

```

	Query Match	6.2%	Score 11.5;	DB 9;	length 780;	
	Best Local Similarity	22.0%;	Pred. No. 10;			
	Matches 33;	Conservative 29;	Mismatches 80;	Indels 15;	Gaps 6;	
QY	10 CNKIEKKMMNMVDVPTVRKRPRILLAAAGSV-ASIKFSNLCH---CFSEMAEYKAIVASKSS	64				
	:   :   :   :   :   :   :   :   :   :					
DQ	282 CENEPEEFENAVNTSCFMPGVLPTDLTVTRLRVKTNKLCEDDDKLMSWSDMSGKKR-	340				
QY	65 LNFVDRPSLPQNVTLTYDDEDEMSS--WNKIGDPYLIELRRMADVMIIAPLSANTLAKTIA	122				
	:   :   :   :   :   :   :   :   :   :					
DQ	341 -NSTGNMKVILQEPICVSIDYSISITCEKMMKNPJTOSTETRLRIYQLVPLLEAHNCIPENN	399				
QY	123 GG---LCDNLITLCIVRANDYSKPLEFVADPAAMNTLMNNDF	158				

```

      ||  :|  | :  ||  :|  | :  ||  | :|  |
Db  400 GGAGCVCHLLMDVVSADNYTLDLW--AGQQLWKGSF 435

```

## RESULT 11

```

US-09-935-868-48
Sequence 48, Application US/09935868
Patent No. US20020164690A1
GENERAL INFORMATION:
APPLICANT: Regeneron Pharmaceuticals, Inc
TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using
FILE REFERENCE: PRT 203D
CURRENT APPLICATION NUMBER: US/09/935,868
CURRENT FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: PCT/US99/22045
PRIOR FILING DATE: 1999-09-22
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.0
SEQ ID NO 48
LENGTH: 782
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-935-868-48

```

## RESULT 12

```

US-09-935-868-52
: Sequence 52, Application US/09935868
: Patent No. US20020164690A1
: GENERAL INFORMATION:
: APPLICANT: Regeneron Pharmaceuticals, Inc
: TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using
: FILE REFERENCE: RRG 203D
: CURRENT APPLICATION NUMBER: US/09/935,868
: CURRENT FILING DATE: 2002-04-11
: PRIOR APPLICATION NUMBER: PCT/US99/22045
: PRIOR FILING DATE: 1999-09-22
: NUMBER OF SEQ ID NOS: 52
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 52
: LENGTH: 782
: TYPE: PRT
: ORGANISM: Homo Sapiens
: US-09-935-868-52

```

Query Match 6.2%; Score 71.5; DB 9; Length 782;  
Best Local Similarity 22.0%; Pred. No. 10;  
Matches 35; Conservative 29; Mismatches 80; Indels 15; Gaps 6;

QY 10 CNKLEKKMMNEVDVYTKPRILLASGSV-ASIKESNLCH----CFSPMAEKKAVASKSS 64  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 282 CENEDEFERNVENTSCFNVPGVLPTLNTVRITRVKTNRKCYEDDKLMSWSQEMSGIKRR- 340  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 65 LNFADKSLPQNATLYIDDEDEWSS--NNKIGDPYLHEELRMADVMIIAPLSANTLAKIA 122  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 341 -NSTGNKKNVLDPEPCVSDYMSISFCENKMMPNPNCSDELRIYDVLVFLLSEAHTCIPNN 399

QY 123 GG---LCDNLTCIVRAMDYSKPLFVAPAMNTLMNNPF 158  
Db 400 GGAGCVCHLMDVVSADNITLDM---AGQOLLMKGSF 435

## RESULT 13

US-10-109-886-8  
; Sequence 8, Application US/10109886  
; Patent No. US20020119499A1  
; GENERAL INFORMATION:  
; APPLICANT: TANABE SEIYAKU CO. LTD.  
; APPLICANT: TANIGUCHI, Tomoyasu  
; APPLICANT: MIZUKAMI, Junko  
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND  
; TITLE OF INVENTION: ANTAGONIST TO PPAR  
; FILE REFERENCE: TANIGUCHI-6  
; CURRENT APPLICATION NUMBER: US/10/109, 886  
; CURRENT FILING DATE: 2002-04-01  
; PRIOR APPLICATION NUMBER: 09/514, 247  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/03734  
; PRIOR FILING DATE: 1998-08-24  
; PRIOR APPLICATION NUMBER: JP231084/1997  
; PRIOR FILING DATE: 1997-08-27  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 2441  
; TYPE: PRT  
; ORGANISM: mouse  
US-10-109-886-8

Query Match 6.2%; Score 70.5; DB 12; Length 2441;  
Best Local Similarity 21.1%; Pred. No. 64;  
Matches 54; Conservative 32; Mismatches 91; Indels 79; Gaps 13;

QY 10 CNKIEKKMNEVDYTR-----KPRLLAASSVASI-----KFSNLCH-- 48  
Db 1179 GSKLAIEFVEDEIDPVMSGLGCGRKYEFSPTLCCYGRQLCTIPRDAAYVSQNKRYHFC 1238  
QY 49 --CFSEMAEYKAVASKSSLNFDKPSLPQNVTLTYDEDESSMNKIGDPVLIHELRRMAD 106  
Db 1239 GKCFTF-----IQGENVTLGDDPSQPQ--TTISKQDFE--KKNDTLDEPPVDCKECGR 1289  
QY 107 VM-----IIAPLSANTLAKIAGGLCDNLTCIVRAMDYSKPLFVAPAMNTLMNNPF 157  
Db 1290 KMHQICVLHYDIIMP-----SGFVCDNCLKTKGRPRKKNK--FSAKRLOTTLRGNH 1338  
QY 158 FTER-----HLVLLDELGITLPIPIKKKLACGDIYNGAMAEPSLIYS 199  
Db 1339 LEDRVNKKFLRRONHPEAGEVFVAVVASSDKTVEVKPKMKSRFV---DSGEMSE--SEPPYR 1393  
QY 200 TVRLFWESQARKORDG 215  
Db 1394 TKALF---AFEIDG 1405

## RESULT 14

US-10-109-886-10  
; Sequence 10, Application US/10109886  
; Patent No. US20020119499A1  
; GENERAL INFORMATION:  
; APPLICANT: TANABE SEIYAKU CO. LTD.  
; APPLICANT: TANIGUCHI, Tomoyasu  
; APPLICANT: MIZUKAMI, Junko  
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND  
; TITLE OF INVENTION: ANTAGONIST TO PPAR  
; FILE REFERENCE: TANIGUCHI-6  
; CURRENT APPLICATION NUMBER: US/10/109, 886  
; CURRENT FILING DATE: 2002-04-01  
; PRIOR APPLICATION NUMBER: 09/514, 247  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/03734

; PRIOR FILING DATE: 1998-08-24  
; PRIOR APPLICATION NUMBER: JP231084/1997  
; PRIOR FILING DATE: 1997-08-27  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 2442  
; TYPE: PRT  
; ORGANISM: human  
US-10-109-886-10

Query Match 6.2%; Score 70.5; DB 12; Length 2442;  
Best Local Similarity 21.1%; Pred. No. 64;  
Matches 54; Conservative 32; Mismatches 91; Indels 79; Gaps 13;

QY 10 CNKIEKKMNEVDYTR-----KPRLLAASSVASI-----KFSNLCH-- 48  
Db 1178 GSKLAIEFVEDEIDPVMSGLGCGRKYEFSPTLCCYGRQLCTIPRDAAYVSQNKRYHFC 1237  
QY 49 --CFSEMAEYKAVASKSSLNFDKPSLPQNVTLTYDEDESSMNKIGDPVLIHELRRMAD 106  
Db 1238 EKCFTF-----IQGENVTLGDDPSQPQ--TTISKQDFE--KKNDTLDEPPVDCKECGR 1288  
QY 107 VM-----IIAPLSANTLAKIAGGLCDNLTCIVRAMDYSKPLFVAPAMNTLMNNPF 157  
Db 1289 KMHQICVLHYDIIMP-----SGFVCDNCLKTKGRPRKKNK--FSAKRLOTTLRGNH 1337  
QY 158 FTER-----HLVLLDELGITLPIPIKKKLACGDIYNGAMAEPSLIYS 199  
Db 1338 LEDRVNKKFLRRONHPEAGEVFVAVVASSDKTVEVKPKMKSRFV---DSGEMSE--SEPPYR 1392  
QY 200 TVRLFWESQARKORDG 215  
Db 1393 TKALF---AFEIDG 1404

RESULT 15  
US-09-870-162A-9  
; Sequence 9, Application US/09870162A  
; Patent No. US20020042118A1  
; GENERAL INFORMATION:  
; APPLICANT: Breinig, Sabine  
; APPLICANT: Fuchs, Georg  
; TITLE OF INVENTION: Phenol Induced Proteins of Thauera aromatica  
; FILE REFERENCE: BC1006 US DIV  
; CURRENT APPLICATION NUMBER: US/09/870, 162A  
; CURRENT FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: 09/516914  
; PRIOR FILING DATE: 2000-03-01  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 9  
; LENGTH: 194  
; TYPE: PRT  
; ORGANISM: Thauera aromatica  
US-09-870-162A-9

Query Match 6.1%; Score 70; DB 10; Length 194;  
Best Local Similarity 34.2%; Pred. No. 2.3;  
Matches 25; Conservative 10; Mismatches 30; Indels 8; Gaps 3;

QY 108 MIAPLSANTLAKIAGGLCDNLTCIVRAMDYSKPLFVAPAMNTLMNNPFTEHRLVLL- 166  
Db 80 MIAPCSIKTILSAVANSFNTNLL--IRADVA--LKERKKVLMREPTLHGHRLMT 134  
QY 167 --DELGITLPIPI 177  
Db 135 QATENGAVLLPL 147

Search completed: November 25, 2002, 10:19:30  
Job time : 12 secs







APPLICANT: Schnell, No. 5837485bert  
APPLICANT: Augustin, Johannes  
APPLICANT: Engelke, Germar  
APPLICANT: Rosenstein, Ralf  
APPLICANT: Kaletta, Corlina  
APPLICANT: Klein, Cora  
APPLICANT: Wieland, Bernd  
APPLICANT: Kupke, Thomas  
APPLICANT: Jung, G nther  
APPLICANT: Kellner, Roland  
TITLE OF INVENTION: Biosynthetic Process For The Preparation  
TITLE OF INVENTION: Of Chemical Compounds  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,625  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/876,791  
FILING DATE: 30-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652.0980002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 181 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

APPLICANT: Kempler, Christoph  
APPLICANT: Jung, Gunther  
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides  
TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/645,1938  
FILING DATE: 13-MAY-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652,1540000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 181 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-08-645-1938-24

Query Match 17.7%; Score 202.5; DB 2; Length 181;  
Best Local Similarity 35.5%; Pred. No. 5.8e-17;  
Matches 59; Conservative 29; Mismatches 57; Indels 21; Gaps 8;

QY 31 LLAAGSVASIKFSNLCCHFSF---MAEYKAVASKSSLNFPVDRKSLPQNVLYTDEDEM 86  
DB 6 LCAAT---ASIVINIVINIIYELKQHFDEVNILFSPSSKNFNTDVL---KLFCDF--- 54

QY 87 SSWNKTGDPVL-HIELRWADVMIIAPLSANTLAKIAGGLCDNLTCIYRAMDYSKPLFV 145  
DB 55 NLYDEIKRPLNHIHIVNHEHYIIVPASANTINKIANGICDNLTYCLT-CYOK-LFI 112

QY 146 APAMNTLMNNPETERHLVLDDELGITLIPPIKKK--LACGDYGN 188  
DB 113 FPNMIRMGNPFLOKNIDLLKNDVKYSPDMKSFSSGRYKN 158

RESULT 5  
US-08-822-774-52  
Sequence 52, Application US/08822774  
Patent No. 6183997  
GENERAL INFORMATION:  
APPLICANT: HOGREFE, HOLLY  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESS: David J. Kulik, Evenson, McKeown, Edwards &  
STREET: 1200 G Street, N.W. Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/822,774  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: KULIK, David J.  
REGISTRATION NUMBER: 36,576  
REFERENCE/DOCKET NUMBER: 1486/43163  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
US-08-822-774-52

Query Match 16.9%; Score 194; DB 4; Length 444;  
Best Local Similarity 31.1%; Pred. No. 2.6e-15;  
Matches 56; Conservative 35; Mismatches 71; Indels 18; Gaps 6;

QY 29 RLLAAGSVASIKFSNLCCHFSF---MAEYKAVASKSSLNFPVDRKSLPQNVLYTDEDEM 87  
DB 31 KIVLGVSGGIAAYKTPELVRRLDRGADVAMTEAKAFITPLST-QAVSGYPVSD--- 86

QY 88 SSWNKTGDPVL-HIELRWADVMIIAPLSANTLAKIAGGLCDNLTCIYRAMD 138  
DB 87 ---SLDDPAEAMGHIETLGGXXKRWADVLVLPATADILIAVAGMADLVSTICLATP 143

QY 139 YSKPLFVAPAMNTLMNNPETERHL-VLDELGITLIPPIKKKACGDYGNAMAPSLI 197  
DB 144 XAAPVAVLPAMNQYRAAATQHNLLEVLAXSRLIHWGPDSSQACGDIIGPERXXDPLTI 203

RESULT 6  
US-09-632-711-52  
Sequence 52, Application US/09632711  
Patent No. 6333165  
GENERAL INFORMATION:  
APPLICANT: HOGREFE, HOLLY  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESS: David J. Kulik, Evenson, McKeown, Edwards &  
STREET: 1200 G Street, N.W. Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/632,711  
FILING DATE: 04-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/822,774  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:

```

NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-09-632-711-52

```

[illegible]

RESULT 7  
 US-09-632-703B-52  
 Sequence 52, Application US/09632703B  
 Patent No. 6379553  
 GENERAL INFORMATION:  
 APPLICANT: HOGREFE, HOLLY  
 TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF  
 Protein Complexes, Isolated PEF Proteins, and Methods for F  
 Identifying Same  
 NUMBER OF SEQUENCES: 61  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.  
 STREET: 1300 I Street, N.W.  
 CITY: Washington  
 STATE: D.C.  
 ZIP: 20005-3315  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: <Unknown>  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/632,703B  
 FILING DATE: 24-Aug-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/822,774  
 FILING DATE: 21-MAR-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BARKER, M. Paul  
 REGISTRATION NUMBER: 32,013  
 REFERENCE/DOCKET NUMBER: 4121.0116-04  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 849-6613  
 TELEFAX: (650) 849-6666  
 INFORMATION FOR SEQ ID NO: 52:  
 SEQUENCE CHARACTERISTICS:

```

?      LENGTH: 444 amino acids
?      TYPE: amino acid
?      TOPOLOGY: unknown
?      MOLECULE TYPE: protein
?      HYPOTHETICAL: NO
?      ANTI-SENSE: NO
?      ORIGINAL SOURCE:
?          ORGANISM: Escherichia coli
?      SEQUENCE DESCRIPTION: SEQ ID NO: 52
?
US-09-632-703b-52

```

	Query Match	16.9%; Score 194; DB 4; Length 444;	
	Best Local Similarity	31.1%; Pred. No. 2.6e-15;	
	Matches	56; Conservative 35; Mismatches 71; Indels 7; Gaps	6
QY	29 RLLAAGSVASIFKSFNLCHEFSEMAEYKA VASKSLNFYDKPSLPONTLYTDEDEMS	87	
	:   :   :   :   :   :   :   :		
Db	31 KIVGSGSIAAYTFPELVRLRDRGCADVRAVMTAKAFIPLLSLQAVSYTPYSD---	86	
QY	88 SWNKIGDPVL-----HIEL---RRADVMITAPLSANTLAKIAGLCDNLITCIVRAND	138	
	:    :     :     :   :   :   :   :   :		
Db	87 ---SLIDPFAAEAMGHIEDLGXXXXKKWADLVILLPATADLIAFAVGAMNDVSTICLATRP	143	
QY	139 YSKPLPYPAANNTLTMMNNPTFERHL-VLLDELGITLIPIPKKKLKACGYDGNAMAEPSSL	197	
	:   :         :   :   :   :   :   :   :   :		
Db	144 XXAPAVALPRANQQYTRAAATQHNIETVLAKSRGLLITMGDSQCAGDGIPGRXXDLPTI	203	

RESULT 8  
 US-09-632-702-52  
 : Sequence 52, Application US/09632702  
 : Patent No. 6444428  
 :  
 GENERAL INFORMATION:  
 : APPLICANT: HOGREEFE, Holly  
 : TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
 : Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
 : and Methods for Purifying and Identifying Same  
 :  
 NUMBER OF SEQUENCES: 61  
 CORRESPONDENCE ADDRESS:  
 : ADDRESSSEE: David J. Kulik, Evenson, McKeown, Edwards &  
 : Lenahan, P.L.L.C.  
 : STREET: 1200 G Street, N.W. Suite 700  
 : CITY: Washington  
 : STATE: D.C.  
 : ZIP: 20005  
 :  
 COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.25  
 :  
 CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/632,702  
 : FILING DATE: 04-Aug-2000  
 : CLASSIFICATION: <Unknown>  
 :  
 PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/832,774  
 : FILING DATE: 21-MAR-1997  
 :  
 ATTORNEY/AGENT INFORMATION:  
 : NAME: KULIK, David J.  
 : REGISTRATION NUMBER: 36,576  
 : REFERENCE/DOCKET NUMBER: 1486/43163  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (202) 628-8800  
 : TELEFAX: (202) 628-8844  
 :  
 INFORMATION FOR SEQ ID NO: 52:  
 :  
 SEQUENCE CHARACTERISTICS:  
 : LENGTH: 444 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: unknown  
 : MOLECULE TYPE: protein  
 : HYPOTHETICAL: NO  
 : ANTI-SENSE: NO  
 :  
 ORIGINAL SOURCE:

```

; ORGANISM: Escherichia coli
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-09-632-702-52

Query Match          16.9%; Score 194; DB 4; Length 444:
Best Local Similarity 31.1%; Pred. No. 2.6e-15;
Matches 56; Conservative 35; Mismatches 71; Indels 18; Gaps 6:

OY 29 RLLAASGSVASIKFSNLCHESE-WAEKAVAKSSSLNFVDPKSLPQNTLYTDEDEMS 87
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 31 KIVLGSGLIAKPELVRLRLRGADRVAMTEAKAFITPLSL-QAVSGTPVSD--- 86
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 88 SMNKICDPVL-----HIEL-----RRADVMIAPLSANTLAKIAGLCDNLTCIYRAMD 138
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 87 ---SLIDPAAEAAMGHIEGXXXXXKMWADLVIIAPATADIIARVAAGMANDLVSTICLATP 143
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 139 YSKPLFVAPAMNTLMNNPFTERRHL-VLDELGITLIPPIKKKLAGDVGNGMAREPSUL 197
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 144 XAPVAVLPAAMQOMYRAAATOHNLVLAIXSRGLTIWPGSGACGDIGPGRXXDPETI 203
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-627-376-9
; Sequence 9, Application US/09627376
; Patent No. 6342385
; GENERAL INFORMATION:
; APPLICANT: Qi, Fengxia          Caulfield, Page          Chen, Ping
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
; FILE REFERENCE: UAB-17402/22
; CURRENT APPLICATION NUMBER: US/09/627,376
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-627-376-9

Query Match          16.5%; Score 188.5; DB 4; Length 188:
Best Local Similarity 28.2%; Pred. No. 3.3e-15;
Matches 42; Conservative 35; Mismatches 63; Indels 9; Gaps 4:

OY 19 MEVDVTRRPRILLASGSVASIKFSNLCHESEMAEYKAVAKSSSLNFVDPKSLPQNV- 77
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MEQNIIEKKILCLTSGSGLGIA-EYITFLYRFKHVRIVYSDNNAKMLPVAITQDCE 59
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 78 TLYTDEDEMSNMKICDPVLHIELRRADVMIAPLSANTLAKIAGLCDNLTCIYRAM 137
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 60 KYTTEDEVSTEDKQK-----NHIALTRWADITVLPATANIGKVANGIADNEMTTLIS- 113
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 138 DYSKPLFVAPAMNTLMNNPFTERRHLVL 166
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 114 -SSKPLVLTFCMNNIMENPVYQKNVEVL 141
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-08-822-774-19
; Sequence 19, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; EXTRACTS, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
US-08-822-774-19

Query Match          14.1%; Score 162; DB 4; Length 403:
Best Local Similarity 27.1%; Pred. No. 2.1e-11;
Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5:

OY 23 TYTRK---PRILLAASGSVASIKFSNLCHESEMAEYKAVAKSSSLNFVDPKSLPQNTLY 79
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11 TKSRLVGRKVIYXXPSIALDVKACEGLIRGAEVHAVMSEAARKIHP----- 61
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 80 YTDEDEMSWN-KIGDPVL-----HIEL-----RRADVMIAPLSANTLAKIAGLC 126
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 -----YAWNLPNGNPVITEITGTEHVELAGEHENEKADLILVCATANTISKIACGID 114
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 127 DNLITCIVRAMDYSKPLFVAPAMNTLMNNPFTERRHLVLDELGITLPIPP 176
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 115 DTVTTVVVTAPRPHIPIMTAPAMHETMYRRPIVRENIERKKRGVFEIG 164
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
US-09-632-711-19
; Sequence 19, Application US/09632711
; Patent No. 6333165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; EXTRACTS, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
```

[illegible]

```

:      HYPOTHEITICAL: NO
:      SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-703B-19

Query Match          14.1%; Score 162; DB 4; Length 403;
Best Local Similarity 27.1%; Pred. No. 2.1e-11;
Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5;

OY   23 TVTRK---PRILLASGSVASIKFSNLCIHCFSRMAEYKAIVAKASSLNIFVVKPSLPQAVTL 79
       |::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
DB    11 TKSRLVKGRIYXXXXGSIADLVKACEGLIRHGAHVAMSEAAFKIIHP----- 61

OY   80 YTDEWSSMN-KIGDPV-----HTEL----RRMADVMIITAPSLANTLAKIAGLC 126
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB    62 -----YANMLPTGNPVTEITETGFIEHYELAGEHENKADLLIVCPATANTISKIACGD 114

OY   127 DNLTCIVRAMDYSKFLFVPAPAMNTLMNNPFTERRLLVLLDELGTILIPP 176
       |::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
DB    115 DPVVTVVTARPHIPIMIAPIAMHETMYRHPIRYRENIERLKIGVERIGP 164

RESULT 13
US-09-632-702-19
: Sequence 19, Application US/09632702
: Patent No. 6444428
: GENERAL INFORMATION:
: APPLICANT: HOGREFE, Holly
: TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
: Extracts, PEF Protein Complexes, Isolated PEF Proteins
: and Methods for Purifying and Identifying Same
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
:           Lenahan, P.L.L.C.
: STREET: 1200 G Street, N.W. Suite 700
: CITY: Washington
: STATE: D.C.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/632,702
: FILING DATE: 04-Aug-2000
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/822,774
: FILING DATE: 21-MAR-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: KULIK, David J.
: REGISTRATION NUMBER: 36,576
: REFERENCE/DOCKET NUMBER: 1486/43163
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 628-8800
: TELEFAX: (202) 628-8844
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 403 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: HYPOTHEITICAL: NO
: SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-702-19

Query Match          14.1%; Score 162; DB 4; Length 403;
Best Local Similarity 27.1%; Pred. No. 2.1e-11;
Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5;

OY   23 TVTRK---PRILLASGSVASIKFSNLCIHCFSRMAEYKAIVAKASSLNIFVVKPSLPQAVTL 79
       |::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|

```

Db 11 TKSRLVGKIVXXXPGSIALDVAKCEGLIRHGAENVHVMSEATKTIHP----- 61  
QY 80 YTDEDESSWN KIGDVL-----HIEI---RMDVMIILAPLSANTLAKIAGLIC 126  
Db 62 -----YAMNLPITGNVTEITGIEHVEELAGEHENKADLLVCPATANTISKIACGID 114  
QY 127 DNLITCIYRANDYSKPLFVAPAMNTLMNNPTEERHVLVLDLGLTLLRP 176  
Db 115 DPTVTVTATAPHPHPIIMAPAMHETMYRHPIVRENIEKLKGVETIP 164

## RESULT 14

US-08-822-774-51  
Sequence 51, Application US/08822774  
Patent No. 6183997  
GENERAL INFORMATION:  
APPLICANT: HOGREFE, HOLLY  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
and Methods for Purifying and Identifying Same  
TITLE OF INVENTION: and Methods for Purifying and Identifying Same  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &  
STREET: 1200 G Street, N.W. Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/822,774  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: KULIK, David J.  
REGISTRATION NUMBER: 36,576  
REFERENCE/DOCKET NUMBER: 1486/43163  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Methanococcus Jannaschii  
US-08-822-774-51

Query Match 11.4%, Score 131; DB 4; Length 437;  
Best Local Similarity 25.8%, Pred. No. 1.6e-07;  
Matches 46; Conservative 33; Mismatches 79; Indels 20; Gaps 4;

QY 3 GTSSFLGCKIEKKMMEVDYTRKPRILLAAAGSVASIKFSNLC-HCFSEMAEYKAVAS 61  
Db 14 GTSKSLLENK-----KILVAVTSSIAAIEPKLRRELRHGAEYVCIT 57  
QY 62 KSSLNFPVKPSLPQNV--TLTYDEDESSWNKISDPVLIHIELRRADYMIILAPLSANTLA 119  
Db 58 EETKTIIGKEALKFCGCGMEVEETIGXXXXXDIHILLYXXXXXNCCDCLLYPATANITS 117  
QY 120 KIAGLIDNLTLCIYRANDYSKPLFVAPAMNTLMNNPTEERHVLVLDLGLTLLRP 176  
Db 118 KINIGIADNIYVTATLMPFGNKPFIIVAPAMHNMENXAXAKRHIDKLEKDKIYIISP 175

## RESULT 15

US-09-632-711-51  
Sequence 51, Application US/09632711  
Patent No. 6333165  
GENERAL INFORMATION:  
APPLICANT: HOGREFE, HOLLY  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
and Methods for Purifying and Identifying Same  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &  
STREET: 1200 G Street, N.W. Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/632,711  
FILING DATE: 04-Aug-2000  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/822,774  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: KULIK, David J.  
REGISTRATION NUMBER: 36,576  
REFERENCE/DOCKET NUMBER: 1486/43163  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Methanococcus Jannaschii  
SEQUENCE DESCRIPTION: SEQ ID NO: 51:  
US-09-632-711-51

Query Match 11.4%, Score 131; DB 4; Length 437;  
Best Local Similarity 25.8%, Pred. No. 1.6e-07;  
Matches 46; Conservative 33; Mismatches 79; Indels 20; Gaps 4;

QY 3 GTSSFLGCKIEKKMMEVDYTRKPRILLAAAGSVASIKFSNLC-HCFSEMAEYKAVAS 61  
Db 14 GTSKSLLENK-----KILVAVTSSIAAIEPKLRRELRHGAEYVCIT 57  
QY 62 KSSLNFPVKPSLPQNV--TLTYDEDESSWNKISDPVLIHIELRRADYMIILAPLSANTLA 119  
Db 58 EETKTIIGKEALKFCGCGMEVEETIGXXXXXDIHILLYXXXXXNCCDCLLYPATANITS 117  
QY 120 KIAGLIDNLTLCIYRANDYSKPLFVAPAMNTLMNNPTEERHVLVLDLGLTLLRP 176  
Db 118 KINIGIADNIYVTATLMPFGNKPFIIVAPAMHNMENXAXAKRHIDKLEKDKIYIISP 175

Search completed: November 25, 2002, 10:16:22  
Job time : 18 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OK protein - protein search, using sw model

Run on: November 25, 2002, 10:10:25 ; Search time 36 Seconds

(without alignments)  
803.205 Million cell updates/sec

Title: US-09-868-300-8

Perfect score: 1145  
Sequence: 1 EFGTSSFLGCKIKFKKMMME.....YSTVRLWESQARKRODCTS 217

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*
- 2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*
- 3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*
- 4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*
- 5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*
- 6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*
- 7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*
- 8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*
- 9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*
- 10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*
- 11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*
- 12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*
- 13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*
- 14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*
- 15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*
- 16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*
- 17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*
- 18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*
- 19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*
- 20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*
- 21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*
- 22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*
- 23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1145	100.0	217	21	AAV96816
2	946	82.6	209	21	AAAG11381
3	943	82.4	199	21	AAAG11382
4	449	39.2	309	22	ABBI1631
5	449	39.2	309	22	AAAB1256
6	445	38.9	204	22	AAAB94179
7	444	38.8	204	22	AAAB94470
8	443	38.7	326	21	AAV32199
9	422	36.9	170	22	AAU18952
10	422	36.9	170	22	AAU17967

11	274.5	24.0	181	23	ABP28494	Streptococcus poly
12	243.5	21.3	178	23	ABBS3882	Lactococcus lactis
13	243.5	21.3	180	23	ABP28493	Streptococcus poly
14	242.5	21.2	399	22	AAAG2741	S. epidermidis ope
15	242.5	21.2	400	23	ABP39940	Staphylococcus epi
16	227.5	19.9	399	23	ABBA4692	Listeria monocytog
17	221	19.3	228	19	AAW98418	H. pylori GHPO 319
18	220	19.2	422	22	AAAB79946	Corynebacterium g1
19	214	18.7	420	22	AAAG91520	C glutamicum prote
20	204.5	17.9	181	14	AAAB93345	Epid protein.. Sta
21	202.5	17.7	181	20	AAV43439	S. epidermis readi
22	195	17.0	269	22	AAU63490	Propionibacterium
23	188.5	16.5	188	23	ABBB09642	Mutd protein of th
24	169	14.8	403	19	AAAB96653	Putative P. abyssi
25	163.5	14.3	401	22	AAAB72844	Polymenase enhanci
26	141	12.3	79	23	ABP08086	Human ORFX protein
27	86	7.5	238	22	ABG07040	Novel human diagno
28	85.5	7.5	669	22	ABAB7461	Amino acid sequenc
29	84.5	7.4	280	22	ABBB0590	Drosophila melanog
30	84.5	7.4	1717	22	ABG20672	Novel human diagno
31	84	7.3	197	22	AAAG91615	C glutamicum prote
32	83.5	7.3	192	20	AAV34858	Chlamydia pneumoni
33	83.5	7.3	505	22	AAAB95727	Human protein sequ
34	82.5	7.2	524	22	ABAB92814	Human protein sequ
35	82	7.2	159	22	AAAG81370	Human AFP protein
36	82	7.2	329	19	AAW98743	H. pylori GHPO 991
37	82	7.2	418	22	AAAG82097	S. epidermidis ope
38	82	7.2	431	23	ABP38206	Staphylococcus epi
39	82	7.2	657	22	AAAB39041	Human polypeptide
40	82	7.2	996	22	ABG30156	Novel human diagno
41	81	7.1	192	20	AAV36979	Protein involved i
42	80.5	7.0	475	22	AAAB93993	Human stomach canc
43	80.5	7.0	475	22	AAAB92735	Human protein sequ
44	80.5	7.0	521	22	ABBS8663	Drosophila melanog
45	80.5	7.0	686	22	AAAB95291	Human protein sequ

ALIGNMENTS

RESULT 1	
AAV96816	
ID	AAV96816 standard; Protein; 217 AA.
XX	
AC	AAV96816;
XX	
DT	26-SEP-2000 (first entry)
XX	
DE	A. thaliana VB89 (HML3) CDC2b interacting protein.
XX	
KW	Cyclin-dependent protein kinase; CDK; CDC2a; CDC2b; VB89; HML3;
KW	cell cycle; interacting protein; environmental stress; growth regulator;
KW	herbicide; nematode resistance; plant breeding.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200036124-A2.
XX	
PD	22-JUN-2000.
XX	
PF	17-DEC-1999; 99WO-EP10084.
XX	
PR	17-DEC-1998; 98EP-0124062.
XX	
PA	(CROP-) CROPESTGN NV.
XX	
PI	De Veylder L, Boudolf VKCK, Torres Acosta JA, Inze D;
XX	
DR	WPI: 2000-431601/37.
XX	
PT	N-PSDB; AAA51412.
XX	
PT	Nucleic acids encoding plant cell cycle interacting proteins, useful
XX	for regulating plant growth and in recombinant DNA protocols

XX Claim 1; Page 126; 152pp; English.  
PS The Vb89 clone encodes the Arabidopsis thaliana HAL3 homologue, a  
XX halotolerant gene isolated in Saccharomyces cerevisiae. The Vb89 clone  
CC interacts with A. thaliana CDC2b (a cyclin-dependent protein kinase  
CC (CDK)), but not with CDC2a in the two-hybrid system.  
CC CDC2a and CDC2b are the only CDK genes to have been characterized in  
CC detail in Arabidopsis thaliana. They were used as bait in a two-hybrid  
CC screening assay with a cDNA library of a plant cell suspension as prey.  
CC The plant cell cycle interacting proteins identified were designated  
CC LDV15, PH80-like protein, Vb33, Vb89, VbDAHP and VbHSF. The nucleic  
CC acids, vectors comprising them, the proteins they express, antibodies  
CC that bind to them and or inhibitors of their protein expression and/or  
CC activity may be used for modulating the cell cycle in an animal or plant,  
CC plant cell division and/or growth, for influencing the activity of cell  
CC cycle proteins in a plant or animal cell, as positive or negative  
CC regulators of cell proliferation, for modifying the growth inhibition  
CC caused by environmental stress conditions (e.g. to improve growth of  
CC plants in normal or suboptimal nutrient conditions, especially  
CC phosphorus), for use in a screening method for inhibitors or activators  
CC of cell cycle protein, as growth regulators, herbicides and/or for  
CC inducing nematode resistance in plants. The DNA sequences and their  
CC regulatory sequences may be used as markers in plant or animal cell and  
CC tissue cultures or as a marker in marker-assisted plant breeding. The  
CC regulatory sequences may also be used for the expression of heterologous  
CC DNA sequences during a stage of the cell cycle.  
SQ Sequence 217 AA:  
Query Match 100.0%; Score 1145; DB 21; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.6e-120;  
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EFGTSSFLCCKNTEKKMNEVDVTTRKPRLLAAGSVASIKRSNCHCFSEMAEKAVA 60  
DB 1 EFGTSSFLCCKNTEKKMNEVDVTTRKPRLLAAGSVASIKRSNCHCFSEMAEKAVA 60  
QY 61 SKSSLNFVDPKSLPONVLTLYTDEDEMSWNKIGDPVLIHELRRADVMIAPLSANTLAK 120  
DB 61 SKSSLNFVDPKSLPONVLTLYTDEDEMSWNKIGDPVLIHELRRADVMIAPLSANTLAK 120  
QY 121 IAGGLCDNLITCIVRAMDYSKPLFVAPAMNTLMNNPTEFRLHVLDELGITLIPRIKK 180  
DB 121 IAGGLCDNLITCIVRAMDYSKPLFVAPAMNTLMNNPTEFRLHVLDELGITLIPRIKK 180  
QY 181 LACGDYGNGAMAPSLIYSTVRLFWESQARKORDGS 217  
DB 181 LACGDYGNGAMAPSLIYSTVRLFWESQARKORDGS 217  
RESULT 2  
AAG11381  
ID AAG11381 standard: Protein; 209 AA.  
XX AAG11381;  
AC AAG11381;  
DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10076.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
XX

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.

PR	13-JUL-1999	9905-01435342
PR	14-JUL-1999	9905-01434605
PR	15-JUL-1999	9905-01440025
PR	16-JUL-1999	9905-01440865
PR	16-JUL-1999	9905-01440866
PR	19-JUL-1999	9905-01443425
PR	19-JUL-1999	9905-01443321
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443334
PR	19-JUL-1999	9905-01443352
PR	20-JUL-1999	9905-01443522
PR	20-JUL-1999	9905-01445376
PR	20-JUL-1999	9905-01445376
PR	21-JUL-1999	9905-01445313
PR	21-JUL-1999	9905-01450866
PR	21-JUL-1999	9905-01450866
PR	22-JUL-1999	9905-01450885
PR	22-JUL-1999	9905-01450885
PR	22-JUL-1999	9905-01450889
PR	22-JUL-1999	9905-01451922
PR	22-JUL-1999	9905-01451945
PR	22-JUL-1999	9905-01451945
PR	23-JUL-1999	9905-01455218
PR	23-JUL-1999	9905-01455218
PR	23-JUL-1999	9905-01455244
PR	26-JUL-1999	9905-01453276
PR	27-JUL-1999	9905-01455113
PR	27-JUL-1999	9905-01455113
PR	27-JUL-1999	9905-01455119
PR	28-JUL-1999	9905-01455951
PR	02-AUG-1999	9905-01463866
PR	02-AUG-1999	9905-01463888
PR	02-AUG-1999	9905-01465889
PR	02-AUG-1999	9905-01465889
PR	03-AUG-1999	9905-01470038
PR	04-AUG-1999	9905-01472044
PR	04-AUG-1999	9905-01473932
PR	04-AUG-1999	9905-01473932
PR	05-AUG-1999	9905-01472600
PR	05-AUG-1999	9905-01473703
PR	06-AUG-1999	9905-01474116
PR	06-AUG-1999	9905-01474163
PR	09-AUG-1999	9905-01478953
PR	09-AUG-1999	9905-01479355
PR	10-AUG-1999	9905-01481711
PR	11-AUG-1999	9905-01483119
PR	12-AUG-1999	9905-01483411
PR	13-AUG-1999	9905-01485655
PR	13-AUG-1999	9905-01486884
PR	16-AUG-1999	9905-01493688
PR	17-AUG-1999	9905-01491735
PR	18-AUG-1999	9905-01494626
PR	20-AUG-1999	9905-01497222
PR	20-AUG-1999	9905-01497233
PR	20-AUG-1999	9905-01499229
PR	23-AUG-1999	9905-01499022
PR	23-AUG-1999	9905-01499330
PR	26-AUG-1999	9905-01500884
PR	27-AUG-1999	9905-01510665
PR	27-AUG-1999	9905-01510665
PR	27-AUG-1999	9905-01510880
PR	30-AUG-1999	9905-01513033
PR	30-AUG-1999	9905-01513388
PR	01-SEP-1999	9905-01519330
PR	07-SEP-1999	9905-01523633
PR	10-SEP-1999	9905-01530758
PR	13-SEP-1999	9905-01537580
PR	15-SEP-1999	9905-01540118
PR	16-SEP-1999	9905-01540779
PR	20-SEP-1999	9905-01545479
PR	22-SEP-1999	9905-01551339
PR	22-SEP-1999	9905-01554866
PR	24-SEP-1999	9905-01536859
PR	28-SEP-1999	9905-01564588
PR	29-SEP-1999	9905-01565969

PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158222
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159283
PR	13-OCT-1999	9905-0159284
PR	13-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159931
PR	14-OCT-1999	9905-0159933
PR	14-OCT-1999	9905-0159637
PR	14-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160885
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160989
PR	25-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161820
PR	28-OCT-1999	9905-0161822
PR	28-OCT-1999	9905-0161943
PR	28-OCT-1999	9905-0162142
PR	29-OCT-1999	9905-0162143

Query Match	82.6%	Score 946;	DB 21;	Length 209;
Best Local Similarity	85.6%	Pred. NO. 4.1e-98;		
Matches 173; Conservative	17;	Mismatches 12;	Indels 0;	Gaps 0;

Oy	16	KMMEVDVTVTKRPRIILLAGSGSVASIKRNSLNCHCSEAEKKAASKSLSNFVKPSLPQ	75
	:	: : : :   : : : : : : : : : :   : : : : : : : : : :   : : : : : : : : : :	
Dd	8	RODMEEVNNTTPPKPRPYLLAASSGVAAIKRGNCICHCEFEWAERAVVTXSSLHFLDKLSLPQ	67
Oy	76	NVTLTVDDEWSSNMKKIGDPVLHTELRRADVMIIAPLASNTLAKIAGGLCDNLTLCIVR	135
		: : : : : : : : : :   : : : : : : : : : :   : : : : : : : : : :   : : : : : : : : : :	
Dd	68	EYTLITDDEWSSNMKKIGDPVLHTELRMDVLYIAPIASNTLTKIAGGLCDNLTLCTIR	127
Oy	136	AMDYSKPFLFVAPAMNTLMWNPNFTERRHLVLLDELGITLIPPIKRKLACGDYGNGAMAEPSS	195
		: : : : : : : : : :   : : : : : : : : : :   : : : : : : : : : :   : : : : : : : : : :	
Dd	128	AMDYTKPLFVAPAMNTLMWNPNFTERRHLSLDELGITLIPPIKRKLACGDYGNGAMAEPSS	187
Oy	196	LITSYRLFWESQARKORDGTS	217
		: : : : : : : : : :   : : : : : : : : : :   : : : : : : : : : :   : : : : : : : : : :	
Dd	188	LITSYRLFWESOAHQDTGTS	209
 RESULT 3 AAG11382 ID AAG11382 standard; Protein; 199 AA. XX AAG11382; XX DT 17-OCT-2000 (first entry) XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 10077. XX KW Protein identification; signal transduction pathway; metabolic pathway; KM hybridisation assay; genetic mapping; gene expression control; promoter KW termination sequence. OS Arabidopsis thaliana. XX XX EPI033405-A2.			

XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 11-MAY-1999; 990S-0132863.  
PR 14-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 17-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.  
PR 30-AUG-1999; 990S-0151303.  
PR 31-AUG-1999; 990S-0151438.  
PR 01-SEP-1999; 990S-0151930.  
PR 07-SEP-1999; 990S-0152363.  
PR 10-SEP-1999; 990S-0153070.  
PR 13-SEP-1999; 990S-0153758.  
PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.  
PR 20-SEP-1999; 990S-0154779.

```
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 82.4%; Score 943; DB 21; Length 199;  
Best Local Similarity 86.9%; Pred. No. 8.4e-98;  
Matches 173; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

```
QY 19 MEVDVTRKPRILLAASSVASIKFNSLCHCFSEMAEYKAVASKSLNFVDKSPQWNT 78
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MEVNTTPKPRVLLAASSVAIKFNGNLCHECTEMAERAVVTKSLHFLDKLSLPEVT 60
QY 79 LYTDEDESSMNKICDPVLIHELRRMADVMITAPLSANTLAKIAGGLCDNLITCIVRAMD 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 LYTDEDESSMNKICDPVLIHELRRMADVLVAPLSANTLAKIAGGLCDNLITCIIIRAMD 120
QY 139 YSKPLFVAPAMTLMWNNPTEMRHVLDELGITLIPRKKLACGDVNGNMAEPLSLTY 198
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 YTKPLFVAPAMTLMWNNPTEMRHLLSLDELGITLIPRKKLACGDVNGMAEPLSLTY 180
QY 199 STVRLFWESQARKORDGTS 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 STVRLFWESQAHQORTGTS 199
```

RESULT 4  
ABBI1631 standard; peptide: 309 AA.  
ID ABBI1631  
XX AC ABBI1631;  
XX DT 11-JAN-2002 (first entry)  
XX DE Human receptor molecule homologue, SEQ ID NO:2001.  
XX KW Human cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;

inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
myeloid cell disorder; lymphoid cell disorder; ashma; arthritis;  
chronic inflammatory condition; proliferative retinopathy;  
atherosclerosis; coronary heart disease; arterial ischaemia;  
bone disorder; osteoporosis; vascular growth disorder;  
tissue regeneration; wound healing; infection; immune disorder;  
cell culture; drug screening; gene therapy; antiinflammatory;  
antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
antifungal; vulnerary; ant ulcer.

Homo sapiens.  
WO200157188-A2.  
09-AUG-2001.  
05-FEB-2001; 2001WO-US03800.  
03-FEB-2000; 2000US-0496914.  
27-APR-2000; 2000US-0560875.  
(HYSE-) HYSEQ INC.  
Tang YF, Liu C, Drmanac RT;  
WPI: 2001-457740/49.  
N-PSDB; ABA08875.  
Human proteins and DNA encoding sequences useful for preventing,  
treating or ameliorating a medical condition in a mammalian subject  
e.g. arthritis and cancer -  
Claim 20; Page 222; 1963pp; English.

Sequences ABBI0981-ABBI2330 represent 1350 novel human polypeptides, and  
sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
invention also relates to vectors and recombinant host cells comprising a  
nucleotide of the invention, methods of producing the novel polypeptides,  
antibodies against the polypeptides, methods of detecting the nucleotides  
or polypeptides in a sample, and methods of identifying compounds which  
bind to polypeptides of the invention. Although novel, many of the  
polypeptides of the invention have homology to known proteins, thereby  
giving an insight into their probable biological activities, and hence  
potential therapeutic applications. The polypeptides of the invention may  
have various activities, including cytokine, cell proliferation or cell  
differentiation activities; stem cell growth factor activity;  
haematopoiesis regulatory activity; tissue growth activity;  
immunomodulatory activity; activin- or inhibin-related activities;  
chemotactic or chemokinetic activities; haemostatic, thrombotic or  
thrombolytic activities; receptor or ligand activities; or may be  
involved in oncogenesis, cancer cell proliferation or metastasis.  
Depending on their biological activities, polypeptides and nucleotides of  
the invention are useful for preventing, treating or ameliorating medical  
conditions, e.g., by protein or gene therapy. Such conditions include  
cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
proliferative retinopathy, atherosclerosis, coronary heart disease,  
arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
vascular growth. Polypeptides involved with tissue regeneration and  
repair (or nucleic acids encoding them) may be used to promote wound  
healing (e.g., of burns, incisions and ulcers), while those with  
immunomodulatory activities may be used in the treatment of viral,  
bacterial and fungal infections in addition to immune disorders.  
Polypeptides with growth factor activity may be used in cell cultures to  
promote cell growth. For example, such polypeptides may be used to  
manipulate stem cells in culture to give rise to neuroepithelial cells  
that can be used to augment or replace cells damaged by illness,  
autoimmune disease or accidental damage. The polypeptides and nucleotides  
may also be used in the diagnosis of the above conditions, and in drug  
screening techniques. The present sequence represents a novel human  
polypeptide of the invention.

```
XX Sequence 309 AA:
SQ
Query Match 39.2%; Score 449; DB 22; Length 309;
Best Local Similarity 46.9%; Pred. No. 7.1e-42;
Matches 92; Conservative 28; Mismatches 70; Indels 6; Gaps 3;

OY 26 RKRPRLLAASGSVASIKFENLCHCFSE--WAQYKAVASKSLSNFVDRKPSLPQNTLYTDE 83
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
DB 51 RKFHVLVGVTSVAALKLPLVSKLIDIPGLEVAVVTTERAKHFYSPQDIP--VTLYSDA 108
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
OY 84 DEWSSWNKIGDPVLIHLELRMDVMIAPLSANTLAKIAGLCDNLITCIVRAMDYSKPL 143
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
DB 109 DEWEMMKSRSDPYLIHIDLRMDLVLAPLDANTLGKVASGICDNLITCIVRAMDRSKPL 168
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
OY 144 FVAPAMNTLMNNPFTERRHLVLDELGITLPIPIKRLACGDYGNGAMAEPSLIYSTVR- 202
   ||||| || : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169 LFCPAMNTAMMEHPITAAQVDQLKAFGYEIPCVAKKLVCGDEGLGAMAEVGTIYDKVE 228
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
OY 203 -LFWESQARKORDGTS 217
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
DB 229 VLFQHSFGFOQSKPGIS 244

RESULT 5
AAM41256
ID AAM41256 standard; Protein; 309 AA.
AC AAM41256;
XX
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 6187.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR
XX 25-APR-2000; 2000US-0552317.
PR
XX 09-JUL-2000; 2000US-0598042.
PR
XX 19-JUL-2000; 2000US-0620312.
PR
XX 03-AUG-2000; 2000US-0653450.
PR
XX 14-SEP-2000; 2000US-0662191.
PR
XX 19-OCT-2000; 2000US-0693036.
PR
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QH, Zhou P, Goodrich R, Dzmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB: AA160412.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2: SEQ ID NO 6187; 10078pp: English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX
XX
```

```
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX SQ Sequence 309 AA:
XX
XX Query Match 39.2%; Score 449; DB 22; Length 309;
XX Best Local Similarity 46.9%; Pred. No. 7.1e-42;
XX Matches 92; Conservative 28; Mismatches 70; Indels 6; Gaps 3;

OY 26 RKRPRLLAASGSVASIKFENLCHCFSE--WAQYKAVASKSLSNFVDRKPSLPQNTLYTDE 83
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
DB 51 RKFHVLVGVTSVAALKLPLVSKLIDIPGLEVAVVTTERAKHFYSPQDIP--VTLYSDA 108
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
OY 84 DEWSSWNKIGDPVLIHLELRMDVMIAPLSANTLAKIAGLCDNLITCIVRAMDYSKPL 143
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
DB 109 DEWEMMKSRSDPYLIHIDLRMDLVLAPLDANTLGKVASGICDNLITCIVRAMDRSKPL 168
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
OY 144 FVAPAMNTLMNNPFTERRHLVLDELGITLPIPIKRLACGDYGNGAMAEPSLIYSTVR- 202
   ||||| || : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169 LFCPAMNTAMMEHPITAAQVDQLKAFGYEIPCVAKKLVCGDEGLGAMAEVGTIYDKVE 228
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
OY 203 -LFWESQARKORDGTS 217
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
DB 229 VLFQHSFGFOQSKPGIS 244

RESULT 6
AAB94179
ID AAB94179 standard; Protein; 204 AA.
AC AAB94179;
XX
XX 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:14492.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
OS
XX
XX EPI074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX
XX 29-JUL-1999; 99JP-0248036.
PR
XX 27-AUG-1999; 99JP-0300253.
PR
XX 11-JAN-2000; 2000JP-0118776.
PR
XX 02-MAY-2000; 2000JP-0183767.
PR
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX
XX
```



DT 15-FEB-2000 (first entry)  
XX Human receptor molecule (REC) encoded by Incyte clone 2022379.  
DE  
XX Receptor; REC; human; diagnosis; therapy; neoplastic disorder;  
KM Immunological disorder; reproductive disorder; nervous disorder;  
KM gastrointestinal disorder; smooth muscle disorder;  
KM musculoskeletal disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO957270-A2.  
XX  
PD 11-NOV-1999.  
XX  
PE 28-APR-1999; 99WO-US09191.  
XX  
PR 01-MAY-1998; 98US-0071822.  
XX  
PA (INCY-) INCYTE PHARM INC.  
PI Hillman JL, Bandman O, Tang YT, Yue H, Lal P, Corley NC;  
PI Guejler KJ, Patterson C;  
XX  
DR MPI; 2000-052971/04.  
XX  
PS N-PSDB; AAZ34608.  
XX  
PT Novel human receptor molecules used in the diagnosis, treatment and  
PT prevention of neoplastic, immunological, reproductive gastrointestinal,  
PT nervous, smooth muscle and musculoskeletal disorders.  
XX  
XX Claim 1; Page 71-72; 94pp; English.  
XX  
XX The present sequence represents a human receptor molecule (REC)  
CC encoded by Incyte cDNA clone 2022379. The invention provides human  
CC RECs and polynucleotides which identify and encode REC, as well as  
CC vectors, host cells, antibodies, agonists and antagonists. Human  
CC RECs appear to play a role in neoplastic, immunological,  
CC reproductive gastrointestinal, nervous, smooth muscle and  
CC musculoskeletal disorders. The protein, antagonists and agonists,  
CC and compositions can be used to treat: a reproductive disorder,  
CC including but not limited to, prolactin production disorders,  
CC infertility including tubal disease, ovulatory defects, endometriosis,  
CC disruptions of the oestrous and menstrual cycles, polycystic ovary  
CC syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian  
CC tumours, uterine fibroids, autoimmune disorders, ectopic pregnancy,  
CC teratogenesis, breast cancer, fibrocystic breast disease,  
CC galactorrhoea, disruptions of spermatogenesis, abnormal sperm  
CC physiology, testis cancer, prostate cancer, benign prostatic  
CC hyperplasia, prostatitis, Peyronie's disease, male breast carcinoma  
CC and gynaecomastia; gastrointestinal disorders including, but are not  
CC limited to, dysphagia, peptic oesophagitis, oesophageal spasm and  
CC stricture, oesophageal carcinoma, dyspepsia, indigestion, gastritis,  
CC gastric carcinoma, anorexia, nausea, emesis, gastroparesis, intestinal  
CC tract infection, peptic ulcer, colitis, Whipple's disease, Mallory-Weiss  
CC syndrome, irritable bowel syndrome, short bowel syndrome, diarrhoea,  
CC constipation, cirrhosis, jaundice, and hepatic vein thrombosis; nervous  
CC disorders including, but are not limited to, Alzheimer's disease,  
CC amnesia, bipolar disorder, cataplexy, cerebral neoplasms, Down's  
CC syndrome, and dystonias; smooth muscle cell disorders including, but  
CC not limited to, angina, anaphylactic shock, arrhythmia, cardiovascular  
CC shock, migraine, and pheochromocytoma; musculoskeletal disorders  
CC including muscular dystrophy, central core disease, nemaline myopathy,  
CC centronuclear myopathy, lipid myopathy, inclusion body myositis,  
CC thyrotoxic myopathy, and ethanol myopathy; immunological disorders  
CC including AIDS, Addison's disease, adult respiratory distress  
CC syndrome, allergy, ankylosing spondylitis, amyloidosis, anaemia,  
CC asthma, atherosclerosis, autoimmune hemolytic anaemia, autoimmune  
CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's  
CC disease, atopic dermatitis, dermatomyositis, diabetes mellitus,  
CC emphysema, episodic lymphopenia with lymphocytotoxic, erythroblastosis  
CC fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,  
CC Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis,  
CC

CC hyperesoinophilia, irritable bowel syndrome, multiple sclerosis,  
CC myasthenia gravis, myocardial or pericardial inflammation,  
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,  
CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's  
CC syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic  
CC sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis,  
CC Werner syndrome, complications of cancer, haemodialysis, and  
CC extracorporeal circulation; viral, fungal, bacterial or protozoan  
CC infections; trauma; and neoplastic disorders including adenocarcinoma,  
CC leukemia, melanoma, myeloma, sarcoma, and various cancers. The REC  
CC polynucleotide is a source of probes and primers which bind may be used  
CC to detect REC in a sample from a patient (claimed). They may also be  
CC administered as part of a gene therapy regime.  
XX  
SQ Sequence 326 AA:  
XX  
XX Query Match 38.7%; Score 443; DB 21; Length 326;  
XX Best local Similarity 48.6%; Pred. No. 3,6e-41;  
XX Matches 87; Conservative 26; Mismatches 62; Indels 4; Gaps 2;  
XX  
QY 26 RKPRLLAAGSGVASIKFNSLCHCFSE--MAEYKAVASKSLNFVDKPSLPQNVTLTYDE 83  
DB 16 RRFHVLVGVTSVAALKPLPLVLSKLDIDPGLEVAVVTTERRAHFTSPQDIP--VILYSDA 73  
QY 84 DEWSWNKIGDPYLHIELRRMADVMIIAPLSANTLAKIAGLCNDLITCIVRAMDYSKPL 143  
DB 74 DEWEWKSRSDPYLHIDLRMADLLVAPLDNLTGKVASGICDNLITCVMAWMDSKPL 133  
QY 144 EVAPAMNTLMNNPFTERRHLVLDELGITLPIPKKLLACGDYGNGAMAPSLISTVR 202  
DB 134 LFCPAMNTAMWEHPITPAQVDOLKAFGYVEIPCVAKKLVCGDEGLGMAEVEGTIVDKVK 192  
XX  
XX RESULT 9  
XX AA018952  
XX ID AAU18952 standard; Protein; 170 AA.  
XX AC AAU18952;  
XX XX  
XX DT 21-NOV-2001 (first entry)  
XX  
XX DE Novel lung cancer antigen, Seq ID No 64.  
XX XX  
XX Human; lung cancer; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytotoxic; cardiact; vasotropic; cerebroprotective;  
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;  
KW ophthalmological; autoimmune disease; rheumatoid arthritis; neoplasia;  
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; skin aging;  
KW ocular disorder; wound healing; organ transplantation.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200155300-A2.  
XX  
XX PD 02-AUG-2001.  
XX  
XX PF 17-JAN-2001; 2001WO-US01238.  
XX  
XX PR 31-JAN-2000; 2000US-0179065.  
XX  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX DR MPI; 2001-465565/50.  
XX  
XX PS N-PSDB; AAS30612.  
XX  
XX PT Isolated nucleic acid molecule encoding a lung cancer antigen is used  
XX in preventing, treating or ameliorating a medical condition  
XX  
XX Claim 1; SEQ ID No 64; 475pp; English.  
XX





[illegible]

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

DR WPI; 2001-476224/51.  
DR N-PSDB; AAS28151.

PT isolated polypeptide for treating, preventing and/ or prognosing  
PT disorders related to the respiratory system including respiratory  
PT cancers and also for testing and detection e.g. diagnosis -  
XS  
Claim 11: SED ID NO 585; 546bp; English.  
PS

Claim 11; SED ID NO 585; 546pp; English.

CC The present invention relates to the isolation of novel human  
CC respiratory antigens, and cDNA (AAS27869-AAS28159) and genomic  
CC sequences encoding for these polypeptides. The sequences of the

invention are useful for preventing, treating and/or prognosing disorders related to the respiratory system including throat disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis), lung disorders e.g. pneumonia, allergic disorders e.g. asthma, pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of the respiratory tissues e.g. lung cancer. The polynucleotide sequences of the invention are useful in gene therapy and antisense therapy. AN017685-AN017975 represent novel human respiratory antigens. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 170 AA;

Query Match	36.9%	Score 422	DB 22	Length 170
-------------	-------	-----------	-------	------------

Best Local Similarity 52.3%; Pred. No. 3.5e-39;  
Matches 81; Conservative 20; Mismatches 50; Indels 4; Gaps 2

```

QY      55 EVKAVASKSSLNFDVKPSLPQNVTLTYDEDEWSSMNNKIGDPVLHIELRRWADYMITAPLS 11
      || :: : : || :: ||| : || | ||||| : ||||| : :: |||
Db     13 EVAVVTTERAKHFYSPODIP--VTLSADAEWEMMKSRSDPVLHIDLRWADLLVAPLD 70

```

QY 115 ANTLAKIAGGLCDNLLTCIVRAWDYSKPLFVAPAMNTLMWNPFETERHLVLLDELGITLI 174

QY 175 PPIKKLACGDYNGMAEPSLIYSTR--LEWES 207

Db 131 PCVAKKLVCDEGLGMAAEVGTIVDKVKEVLFQHS 165

RESULT 11  
ABP28494

ID	ABP28494	standard; Protein; 181 AA
----	----------	---------------------------

AC ABP28494;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 6164.

KW Streptococcus; GAS, GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus pyogenes.

PN WO200234771-A2

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHTR-) CHIRON SPA.

XX

PI Tettelin H;

DR WPI; 2002-352536/38

XX

PT or disease caused by Streptococcus bacteria, such as

PS Claim 1; Page 3779; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B



XX Claim 1: Page 3778; 4525bp; English.  
PS  
XX  
CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC Streptococcus pyogenes (Streptococcus agalactiae) or group A Streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6004-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.  
CC  
XX  
SQ Sequence 180 AA:  
Query Match 21.3%; Score 243.5; DB 23; Length 180;  
Best Local Similarity 32.6%; Pred. No. 4.9e-19;  
Matches 59; Conservative 43; Mismatches 64; Indels 15; Gaps 4;  
QY 29 RLLAASGSVASIKFSNLCHCFSEMA-EVKAVASKSLNFDKPSL-----PQNTLYT 81  
DB 4 RLLAVTGSISYKKAADLSQLTKIGYDVHITMQATEFIRPLTVLSKNPIHLDVMD 63  
QY 82 DEDESSWNKIGDPVLIHELRRADVMIIAPLSANTLAKIAGGLCDNLTCIYRANDYSK 141  
DB 64 ENP-----KI---INHIELAKRTDLFIYAPASANTIAHLAAGFANITYTSVALAPDET 115  
QY 142 PLFVAPAMNTLMNNPFTERRHLVLLDELGITLIPPIKKKLACGDYNGAMAEPISLYTV 201  
DB 116 PLFIAPAMNTKMYHNTIRNIDILKKIGYOEIEPRISLACGDYNGALADISTILKCI 175  
QY 202 R 202  
DB 176 Q 176  
RESULT 14  
AAG82741  
ID AAG82741 standard; Protein; 399 AA.  
XX  
AC AAG82741;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2576.  
XX  
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KM vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
PN WO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000WO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAXO ) GLAXO GROUP LTD.  
XX  
PI Kimmery WJ;  
XX  
DR WPI: 2001-316495/33.  
C DR N-PSDB; AAH53591.

XX  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
PS Claim 18; Page 675; 2188bp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only discloses up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 399 AA:  
Query Match 21.2%; Score 242.5; DB 22; Length 399;  
Best Local Similarity 32.8%; Pred. No. 1.9e-18;  
Matches 61; Conservative 34; Mismatches 74; Indels 17; Gaps 4;  
QY 30 ILLAASGSVASIKFSNL-CHCFSEMAEYKAVASKSLNFDKPSL-----PQNTLYT 82  
DB 4 ILLAVTGSIAAAYKAIDUTSKLQSGYDVVWLSDNAQEFVTPLAQAIISRNPVYNTFERE 63  
QY 83 EDDESSWNKIGDPVLIHELRRADVMIIAPLSANTLAKIAGGLCDNLTCIYRANDYSK 142  
DB 64 ENP-----BEIQHVSIGDMADAIVAPATANTIAKISVGIADLITSTILA--TTTP 113  
QY 143 LFVAPAMNTLMNNPFTERRHLVLLDELGITLIPPIKKKLACGDYNGAMAEPISLYTV 202  
DB 114 KEVAPAMNTVMYNNPFTKHNKRVLSQDGYTFTEPGSGYLACGYVAKGRMEPEMQLISVIN 173  
QY 203 LFWESEQ 208  
DB 174 KEFTQQ 179  
RESULT 15  
ABP39940  
ID ABP39940 standard; Protein; 400 AA.  
XX  
AC ABP39940;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4785.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KM antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.

```

XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI: 2002-381255/41.
XX N-PSDB; ABN92485.
XX
PI Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PI polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure; SEQ ID 4785; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 400 AA;
XX
Query Match 21.2%; Score 242.5; DB 23; Length 400;
Best Local Similarity 32.8%; Pred. No. 1.9e-18;
Matches 61; Conservative 34; Mismatches 74; Indels 17; Gaps 4;
OY 30 ILLAAGSVASIKFSNL-CHCFSEMAEYKAVASKSSLNFVDKPSL-----PQNTLYTD 82
Db 5 ILLAVTGGLAAVKAIDLSKLIQSGYDVAVMLSDHAQEFVTPPLAFOATSRNPVYNTFEKE 64
OY 83 EDEMSSWKKIDPVIHIELRRADVMIIAPLSANTLAKIAGLCNLTCTYRANDYSKP 142
Db 65 ENP-----EEIQHVSIGWADAIIVAPANTIAKLSVGIADDLITSTLA--TTTP 114
OY 143 LFVAPAMNTLMNNPFTEPHVLDELGCTILPIPKKLCAGDYGNGAMAEPSLIYSTVR 202
Db 115 KFYAPAMVNNYNNPRTKHNNKVLSDGCTFYEPGSGYLACGYVAKGRMEPMQILSVIN 174
OY 203 LFWESEQ 208
Db 175 KFTQQ 180

```

Search completed: November 25, 2002, 10:14:35  
 Job time : 39 secs

**THIS PAGE BLANK (USPTO)**